

NASA Conference Publication 10160

CENDI Indexing Workshop
NASA Headquarters
September 21-22, 1994

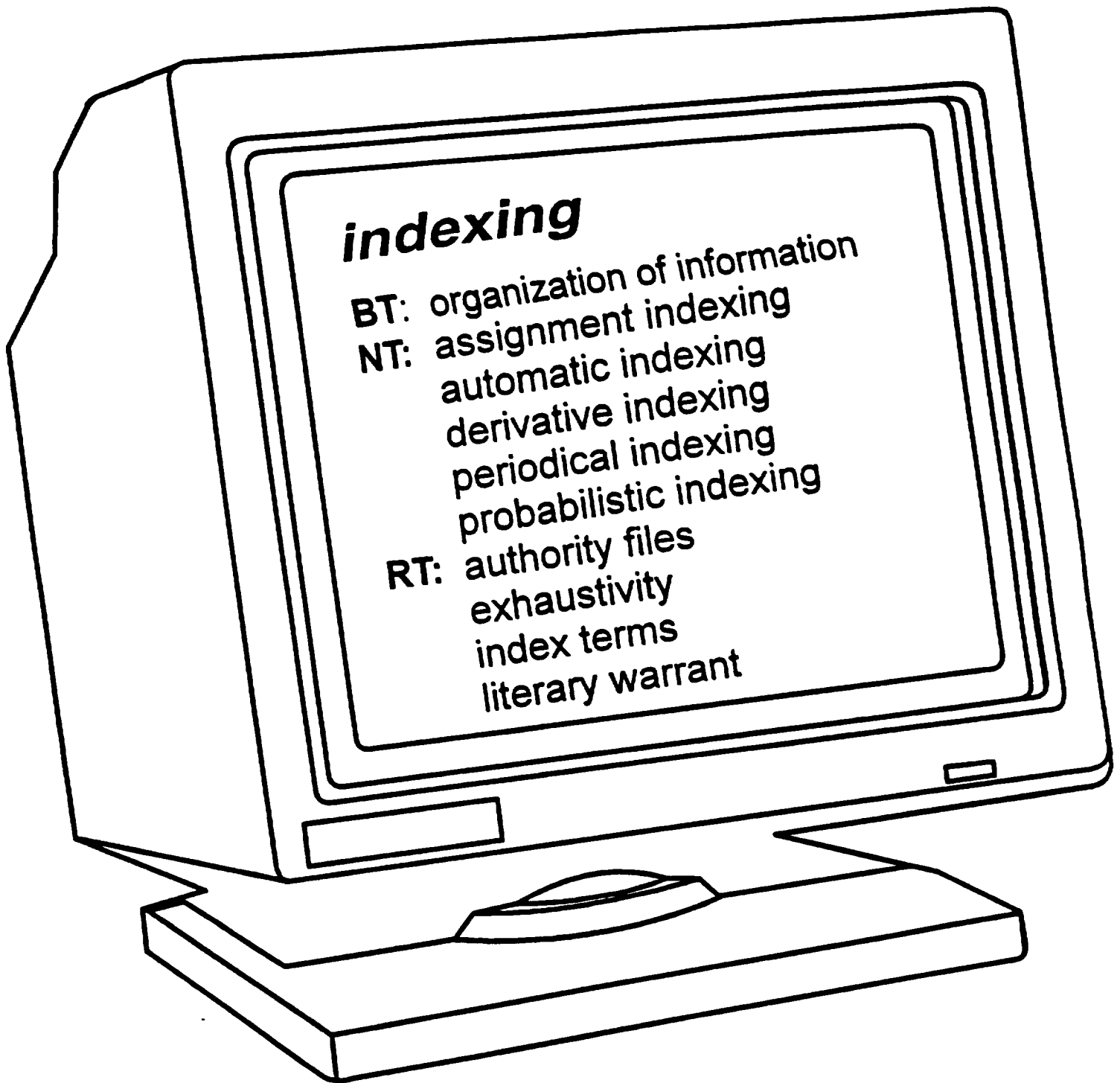


National Aeronautics and Space Administration
Scientific and Technical Information Program
Washington, DC 1994

This publication was prepared by the NASA Center for AeroSpace Information,
800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934, (301) 621-0390.

CENDI Indexing Workshop

September 21 • 22, 1994



**NASA Headquarters
Two Independence Square
300 E Street, SW
Washington, DC**

CENDI Indexing Workshop

September 21 - 22, 1994

CENDI Agencies

Department of Commerce

Department of Energy

**National Aeronautics and
Space Administration**

National Library of Medicine

Department of Defense

**NASA Headquarters
Two Independence Square
300 E Street, SW
Washington, DC**

CENDI INDEXING WORKSHOP

September 21 - 22, 1994

NASA Headquarters ♦ Two Independence Square ♦ 300 "E" Street, SW ♦ Washington, DC

AGENDA

Wednesday, September 21, 1994
8:00 a.m. to 4:00 p.m. Room MIC 5A

Thursday, September 22, 1994
8:00 a.m. to NOON Room MIC 5B

WEDNESDAY MORNING

8:00 - 8:30 **REGISTRATION/ COFFEE**

8:30 - 8:45 Introduction: Focusing the Work Shop *Terese Ohnsorg, NASA STI Program*

8:45 - 9:45 National Air Intelligence Center (NAIC) *Charles Simmons, Donald Quigley*

MAI Project
Enhanced Open-source Service Agent
Application of SGML
"Database It or Retrieve It"

9:45 - 10:15 Defense Technical Information Center (DTIC) *Shirley Kye*

Indexing Quality
Indexing Pilot Project
Unclassified, Unlimited Technical Reports on WAIS

10:15 - 10:30 **BREAK**

10:30 - 11:15 National Library of Medicine (NLM) *Susanne Humphrey*

MedIndEx Prototype

11:15 - 11:45 NASA STI Program *Elliott Linder*

MAI: Maintenance, Performance, Development

11:45 - 12:15 Department of Energy/Office of Scientific and
Technical Information (DOE/ OSTI) *Charles Stuber*

DOE/OSTI Indexing Activities

12:15 - 1:15 **LUNCH**

WEDNESDAY AFTERNOON

1:15 - 1:45	<u>Council on Library Resources</u>	<i>Julia Blixrud</i>
	GCDIS Thesaurus Pilot Project (CONQUEST)	
1:45 - 2:15	<u>Defense Intelligence Agency</u>	<i>George Thompson</i>
	TIPSTER	
2:15 - 2:45	<u>BIOSIS</u>	<i>Colleen Finley</i>
	Category Indexing Schemes	
2:45 - 3:00	BREAK	
3:00 - 3:30	<u>CorpTech</u>	<i>Steven Caldwell</i>
	High-Tech Coding Structures	
3:30 - 4:00	<u>US Geological Survey</u>	<i>Eliot Christian</i>
	Government Information Locator Service	
4:00 - 4:15	SUMMARY & WRAP-UP	

THURSDAY MORNING

CENDI Discussion Leading to Action Items to be Proposed to CENDI Principals

8:30 - 9:00	Listing of Proposed Action Items for Discussion
9:00 - 9:30	DOE/OSTI Proposal to Develop a Category Indexing Scheme
9:30 - 10:00	Application of NLM's MedIndEx Prototype
10:00 - 10:15	BREAK
10:15 - 11:15	Possible CENDI Member Joint Applications of TIPSTER, CorpTech, CRS, and Other Schemes
11:15 - 11:45	Interfacing with CENDI Search Engine Task Force
11:45 - NOON	SUMMARY & WRAP-UP

CENDI Indexing Workshop
NASA Headquarters
September 21 - 22, 1994

Attendees

Steve Caldwell	Corptech 12 Alfred St. Woburn, MA 01801	(617) 932-3939	
Andrew Campbell	Corptech 12 Alfred St. Woburn, MA 01801	(617) 932-3939	
Bonnie Carroll	IIa/CENDI P.O. Box 4219 Oak Ridge, TN 37831	(615) 481-0338	bcarroll @sti.nasa .gov
John H. Dickert	DTIC Rm 5A317 Cameron Station Alexandria, Va 22308	(703) 274-6876	jdickert @dgis. dtic. mil
Beth Duston	IIa/CENDI 1801 Crystal Dr. #905 Arlington, VA 22202	(703) 521-4306	bduston @world. std.com
Colleen Finley	BIOSIS 2100 Arch St. Phila. PA 19003	(215) 587-2003	
Joan Gilbert	DTIC/OCS	(703) 274-6876	
Dug Greevy	NTIS 5285 Port Royal #306 Springfield, VA 22161	(703) 487-4003	dgreevy@ ntis. fedworld. gov
Greg Guthrie	NTIS	(703) 487-4640	
Huddy Haller	DTIC	(703) 617-7931	hhaller @dgis. dtic.dla. mil

Anne Harrison	NTIS 5285 Port Royal Forbes 305 Springfield, VA 22161	(703) 487-4643	
Barbara Henry	NASA (CASI)	(301) 621-0116	bhenry@ sti.nasa. gov
Susanne Humphrey	Nat.Lib. of Medicine Bethesda, MD 20894	(301) 496-9300	humphrey@ nlm.nih. gov
Carol Jacobson	DTIC-E	(703) 274-5367	jacobson@ dgis.dtic dla.mil
James E. Kent	NAIC/WPAFB	(513) 257-6219	
Shirley Kye	DTIC	(703) 274-5367	skye@dgis dtic.dla. mil
Elliott Linder	NASA STI/ IIA	(703) 416- 5624	elinder @sti.nasa gov
Teresa McGervey	NTIS	(703) 487-4640	
Myra McLennan	DTIC-OCS	(703) 274-6876	
Jessica Milstead	JELEM Co. P.O. 5063 Brookfield, CT 06804	(203) 740-2433	76440. 2356@ compuser .com
Anh Nguyen	DTIC/OCS	(703) 274-6876	
Terese Ohnsorg	NASA HQ Code JTT 4th & E Sts. SW Washington, DC	(202) 359-1386	tohnsorg@ hqops. nasa.gov
Don Quigley	NAIC/SCX/SPAFB	(513) 257-4671	
Carol Rouch	Battelle 10-305b 505 King Ave. Columbus, OH 43201	(614) 424-6192	

Kevin Ryan	DTIC-OCS	(703) 274-6876	
Jim Schroer	NASA -CASI	(301) 621-0154	jschroer @sti.nasa .gov
June Silvester	NASA-CASI	(301) 621-0118	jsilveste r@sti. nasa.gov
Chuck Simmons	NAIC/WPAFB	(513) 257-6298	
Roy Stiltner	NASA/CASI	(301) 621-0131	rstiltner @sti. nasa.gov
Charlie Stuber	DOE/OSTI P.O. Box 62 Oak Ridge, TN 37831	(615) 576-1193	charlie. stuber@ ccmail. osti.gov
Annie Washington	DTIC-OCS	(703) 274-5441	
Sue Willis	NTIS	(703) 487-4640	

John Wilson

IIA

(703)
416-5623

TABLE OF CONTENTS

1. National Air Intelligence Center	1
2. Defence Technical Information Center	9
3. National Library of Medicine	21
4. NASA Scientific and Technical Information	103
5. DOE/OSTI Office of Scientific and Technical Information	113
6. Defense Intelligence Agency	127
7. RMS Associates	137
8. CorpTech	175
9. US Geological Survey	191
10. Action Items	247

1

National Air Intelligence Center

NAIC Machine Assisted Indexing (MAI) System

Phase 1 MAI will provide immediate help for the indexer in late 1994

- MAI is a contractor developed, U.S. Government owned client-server software application
- It will allow the indexer to easily format text and indices by means of a user friendly GUI
- It will integrate several of the processes required to create a massive relational database

The future of MAI is promising

- Because the system is being modularly designed, additional capabilities can be easily added
- Automate PFN selection, text markup, and QC
- Incorporate foreign language text processing
- Leverage off other U.S. Government projects - Project Pathfinder, Intelligence Analyst Associate, Machine Translation
- Port MAI to other platforms - Windows, MAC, etc

NAIC is very excited about MAI's potential

- It will allow NAIC to meet expanded mission requirements by taking advantage of technological opportunities
- Phase 1 will give indexers immediate help in dealing with this glut of information
- Future MAI development will facilitate the community's move toward full-text, multi-media databases

NAIC Machine Assisted Indexing (MAI) System

National Air Intelligence Center MACHINE ASSISTED INDEXING SYSTEM

POINTS OF CONTACT:

MR. CHARLES D. SIMMONS/DXLI/(513)257-6298
CAPT JIM KENT/DXLX/(513)257-6219

Why MAI?

- NAIC's mission and information requirements have expanded dramatically
 - NAIC's mission has expanded from the Soviet Union's role in the realm of Science and Technology (S&T) and to the entire world's role in S&T and General Military capabilities
- New Technological Opportunities
 - NAIC is evolving from an IBM mainframe based computing environment to a UNIX, client-server environment

Central Information Reference and Control (CIRC) System

- CIRC contains 10,000,000+ relational records dating back to the 1960's
- The system records contain bibliographic, textual abstracts or extracts, and deeply indexed Personality, Facility, Nomenclature (PFN) data

Current Indexing System requires intensive human labor

- All data elements are created on-line, but require vast amounts of manual keying
- Document preparation, indexing, quality control, and data loading processes are not integrated
- There is little on-line help for the indexer

NAIC Machine Assisted Indexing (MAI) System

Phase 1 MAI will provide immediate help for the indexer in late 1994

- MAI is a contractor developed, U.S. Government owned client-server software application
- It will allow the indexer to easily format text and indices by means of a user friendly GUI
- It will integrate several of the processes required to create a massive relational database

The future of MAI is promising

- Because the system is being modularly designed, additional capabilities can be easily added
- Automate PFN selection, text markup, and QC
- Incorporate foreign language text processing
- Leverage off other U.S. Government projects - Project Pathfinder, Intelligence Analyst Associate, Machine Translation
- Port MAI to other platforms - Windows, MAC, etc

NAIC is very excited about MAI's potential

- It will allow NAIC to meet expanded mission requirements by taking advantage of technological opportunities
- Phase 1 will give indexers immediate help in dealing with this glut of information
- Future MAI development will facilitate the community's move toward full-text, multi-media databases

NAIC Machine Assisted Indexing (MAI) System

National Air Intelligence Center MACHINE ASSISTED INDEXING SYSTEM

POINTS OF CONTACT:

MR. CHARLES D. SIMMONS/DXLL/(513)257-4298
CAPT JIM KENT/DXLL/(513)257-4219

Why MAI?

- NAIC's mission and information requirements have expanded dramatically
 - NAIC's mission has expanded from the Soviet Union's role in the realm of Science and Technology (S&T) and to the entire world's role in S&T and General Military capabilities
- New Technological Opportunities
 - NAIC is evolving from an IBM mainframe based computing environment to a UNIX, client-server environment

Central Information Reference and Control (CIRC) System

- CIRC contains 10,000,000+ relational records dating back to the 1960's
- The system records contain bibliographic, textual abstracts or extracts, and deeply indexed Personality, Facility, Nomenclature (PFN) data

Current Indexing System requires intensive human labor

- All data elements are created on-line, but require vast amounts of manual keying
- Document preparation, indexing, quality control, and data loading processes are not integrated
- There is little on-line help for the indexer

A Proposal For Establishing
Data Standards within the
National Open Source Community

A variety of components necessary for the implementation of a national level Open Source Information System (OSIS) for the Intelligence Community (IC) have previously been spelled out - i.e., an architecture covering connectivity and functionality, security policy, executive agency procedures. IC members who are participants in the OSIS have moved out in a variety of directions to implement pieces of the overall system. One component of the OSIS which has not been addressed in a cohesive fashion is that of data element standards. Currently, each participant is developing data element definitions for their local interests with little regard to the needs of the community as a whole. This approach jeopardizes the community's ability to perform information transfer and, in the end, will confound information understanding by the end customer. As each OSIS participant implements information sources at their respective agencies, and the number of these multiply, the OSIS will find itself mired with unintelligible collections of exclusive, often contradictory data standards. And, the problem is larger than the intelligence community when one reflects on the interactions between the technical literature used with the community and that of other government communities such as the CENDI.

NAIC believes that the time is at hand to set forth an agreed to set of data elements for bibliographic data referencing. Further, these data elements should be inclusive of the needs of the major open source communities within the government and not just those of the intelligence community. At the same time that a standard set of data elements are established, standard SGML tags as well as HTML tags should be developed. NAIC proposes that the IC and CENDI establish a joint data element standards committee that can be used as a springboard for further discussions with other government open source participants. Initially, the committee should limit its consideration to bibliographic data elements for document reference information. Additional areas for standardization can be taken on as each preceding area is completed.

As an initial strawman NAIC can offer the data element standards used in the DOD STIISP and its national data base - CIRC. This proposal is not intended to imply that the CIRC data standards would indeed end up as the community standard, but the CIRC data standards do represent the requirements of a large contingent of the DOD Intelligence Community. We would suggest that each organization bring forth their data standards to the committee with an attitude of cooperation and compromise. NAIC would be glad to chair such a committee or be a member of such a committee.

Defense Technical Information Center

PRECEDING PAGE BLANK NOT FILMED

IMPROVE INDEXING AND RETRIEVAL PROJECT

*DEFENSE TECHNICAL
INFORMATION CENTER
(DTIC)*

Directorate of Information Science & Technology

BACKGROUND

■ Defense RDT&E Online System (DROLS)

- Technical Report (TR) Bibliographic Database
- Work Unit Information System (WUIS)
- Independent Research and Development (IR&D) Database

■ Current indexing methods at DTIC

GOAL

- To improve the quality of DTIC's indexing and retrieval

Directorate of Information Science & Technology

ASSUMPTIONS

- Indexing is a major factor in successful subject retrieval
- As DTIC modernizes, we need to examine the effectiveness of our indexing

OBJECTIVES

- Phase 1: Determine the baseline quality of DTIC's present indexing
- Phase 2: Perform failure analysis and develop an indexing philosophy
- Phase 3: Develop indexing policies and procedures and select indexing software to be used with DTIC's Electronic Document Management System (EDMS)

ACCOMPLISHMENTS

(PHASE 1)

- Selected 19 queries from the DTIC query files
- Searched in TR Database and WUIS
- Processed data and generated reports
- Conducted relevance judging
- Calculated recall and precision ratios for TR Database and WUIS
- Prepared report

CONCLUSIONS OF PHASE 1

- Improvements are needed in indexing and searching
- Customers need to use different capabilities of DROLS for high recall

PROJECT MILESTONES

(PHASE 2)

Perform failure analysis and develop an indexing philosophy (Aug 94 -May 95)

- Conduct failure analysis
- Prepare report on failure analysis
- Review DTIC documents and interview appropriate DTIC/Information Analysis Center (IAC) staff and DTIC customers
- Prepare statement of indexing philosophy

PROJECT MILESTONES

(PHASE 3)

Develop indexing policies and procedures and select indexing software for EDMS

- Evaluate DTIC's Thesaurus and Machine Aided Indexing (MAI)
- Prepare statement of indexing policies
- Prepare general functional specification and system specification for indexing software
- Recommend indexing software package
- Procure software and implement

Directorate of Information Science & Technology

2010年10月10日

3

National Library of Medicine

PRECEDING PAGE BLANK NOT FILMED

MedIndEx™

Project Staff and Collaborators

Susanne M. Humphrey, Project Leader
Lister Hill National Center for Biomedical Communications
National Library of Medicine

Management System Designers, Inc.,
General Software Support Contractor

Index Section and MeSH Section, Collaborators
Library Operations
National Library of Medicine

Organization of talk:

- **conventional database indexing**
- **knowledge-based indexing**
- **frame (object-oriented) data structure**
- **other parts of system**
- **presumed advantages of approach**
- **evaluation**

Conventional Indexing

MeSH Tools (Publications)

MeSH Annotated Alphabetic List

MeSH Tree Structures

Permuted MeSH

Supplementary Chemical Records

Conventional Indexing

MeSH

Preferred terms for unique concepts:

18,000 headings (descriptors)

82 subheadings (topical qualifiers)

78,000 supplementary chemical terms

Official alternate terms

(i.e., synonyms, abbreviations, variants)

Conventional Indexing

MEDLINE Indexing Tools (Publications)

Annotations in alphabetic MeSH

Indexing manual

Technical notes

Specialized manuals (tumors, enzymes, etc.)

**Extensive training, supervised trainee period,
annual orientation meeting, memos, etc.**

Conventional Indexing

Interactive Data Entry

Validation of entered terms

Substitution with preferred terms

Validation of MH/SH pre-coordinations

Relatively simple warnings

User-requested MeSH displays

MEDLINE searching

(simultaneous windows NLM staff only)

Cut & paste (NLM staff only)

Online indexing documentation

(NLM staff future)

General Indexing Rules

Specificity

Index to the most specific term available.

Multiplicity

Provide as many indexing terms as necessary.

Coordination

Express concepts by combining two or more indexing terms.

Heading/Subheading

Attach subheading to the heading which it qualifies.

Coordinate Indexing

bone loss in postmenopausal women
is pre-coordinate heading:

Osteoporosis, Postmenopausal

Coordinate Indexing

***spinal osteoporosis* is coordination of:**

**Osteoporosis
+
Spinal Diseases**

Coordinate Indexing

prevention of bone loss

is heading-subheading pre-coordination:

Osteoporosis /PREVENTION & CONTROL

Coordinate Indexing

estrogen replacement therapy for bone loss
is coordination of headings:

Estrogen Replacement Therapy
+
Osteoporosis

Coordinate Indexing

*estradiol in estrogen replacement therapy
for prevention of spinal bone loss in
postmenopausal women
is coordination of:*

Estrogen Replacement Therapy

+

**Osteoporosis, Postmenopausal /PREVENTION
& CONTROL**

+

Estradiol/THERAPEUTIC USE

+

Spinal Diseases /PREVENTION & CONTROL

MESH ANNOTATED ALPHABETIC LIST

...

BONE LOSS, AGE-RELATED
see **OSTEOPOROSIS**

...

BONE LOSS, PERIMENOPAUSAL
see **OSTEOPOROSIS, POSTMENOPAUSAL**

...

BONE LOSS, POSTMENOPAUSAL
see **OSTEOPOROSIS, POSTMENOPAUSAL**

...

ESTROGEN REPLACEMENT THERAPY
coord IM with specific estrogen /ther use (IM or NIM)
XR MENOPAUSE

...

MENOPAUSE
see related
ESTROGEN REPLACEMENT THERAPY
OSTEOPOROSIS, POSTMENOPAUSAL

...

OSTEOPOROSIS

...

OSTEOPOROSIS, POSTMENOPAUSAL

...

Coordinate Indexing

MeSH TREE STRUCTURES

E2-PROCEDURES AND TECHNIQUES- THERAPEUTIC

THERAPEUTICS	E2
...	...
DRUG THERAPY	E2.319
...	...
ESTROGEN REPLACEMENT THERAPY	E2.319.347
...	...
...	...

MEDLINE Citation

TI -Monitoring skeletal response to estrogen.

AB -Estrogen replacement therapy at accepted doses is not fully effective in preventing bone loss and fractures in postmenopausal women. Bone densitometry is useful for monitoring estrogen replacement therapy to assess dose, foster compliance, and check for secondary bone loss. The most appropriate site for bone loss monitoring is probably the spine because it shows ...

**MH -Absorptiometry, Photon
Bone Density /*DRUG EFFECTS
*Estrogen Replacement Therapy
Female
Human
Menopause/PHYSIOLOGY
Monitoring, Physiologic
Tomography, X-Ray Computed**

Coordinate Indexing

Concepts not Indexed

preventing bone loss

spinal osteoporosis

preventing fractures

doses

What MedIndEx is Not

**Not a new set of concepts representing
biomedicine**

(We have MeSH for that)

Not a new indexing scheme

**(We have MEDLINE indexing rules
for that)**

What is new about MedIndEx

**Indexing rules as executable
computer code**

**Indexing concepts subdivided by
explicit domain-specific relations**

**Concepts, with their associated
relations and rules, organized as
an inheritance hierarchy**

Some Features of Conventional Indexing Useful for Knowledge-Based Indexing

Rules as encodable procedural knowledge

Factual knowledge, including classification

Implied relations

Windowing / mousing environment

Coordinate Indexing Encoded as Prompts

Implied relations as prompts

***Alphabetic MeSH* annotation at Lymph Nodes:
“cervical lymph nodes” = Lymph Nodes + Neck**

Lymph Nodes

LOCATION: Neck

***Indexing Manual* rule: Every neoplasm must be indexed with a histologic type term and a term for the anatomic site**

Bone Neoplasms

BODY-SITE: Ribs

HISTOLOGIC-TYPE: Adenocarcinoma

Coordinate Indexing Encoded as Prompts

***MeSH Trees* categories as prompts**

Drug Therapy

PROBLEM (Disease category)

AGENT (Drugs and Chemicals category)

Subheadings as prompts

Testosterone

ADMINISTRATION/DOSAGE

(administration & dosage subheading)

TOXIC-EFFECT (ad-poi-tox subheadings)

MEDINDEX		PRINT FILE ON	
		SCOPE OFF	
		ANALYZE TI/AB OFF	
INITIAL-FRAME - Concept Frame > GENDER nil AGE nil HUMAN/ANIMAL/PREGNANCY/HOMINIDAE nil CASE-REPORT/COMPARATIVE-STUDY/IN-V nil STUDY-METHOD/DESIGN/FACTOR/MEASURE nil PUBLICATION-TYPE nil NON-US-GOV/US-GOV-PHS/US-GOV-NON-P nil	Concept Frame nil	Restrictions Medical History Subject in Action	RESTRICTIONS HIERARCHY SCOPE NOTE EXPLAIN SLOT
	INITIAL-FRAME - Concept Slot INITIAL-MAJOR-TOPIC > bone loss	Bone Resorption Osteoporosis Osteoporosis, Postmenopausal	INDICES MEDLINE INDICES SAVED MESSAGES CONSIST CHECK SUBJECT FRAME DOCUMENT FRAME CAN/RESTORE FR RESTART DOC CANCEL DOC SEND MAIL
Active Topics Bone Resorption / Osteoclastic Bone Loss Osteoporosis / Age-Related Bone Loss Osteoporosis, Postmenopausal / Perimenopausal Bone Loss ; Postmenopausal Bone Loss			
Enter one or An initial to substance, wh			

Restrictions: Duplex

Estrogens

- . Estradiol / 17 beta-Estradiol ; Estradiol-17 beta ;
- . Estriol
- . Estrogens, Catechol
- . Estrogens, Conjugated / Premarin ;
- . Estrogens, Non-Steroidal
- . Estrone
- Estrogens, Synthetic / Estrogen Analogs ;
- . Estrogens, Non-Steroidal
- . . Diethylstilbestrol / Stilbestrol ;
- . . Ethinyl Estradiol / Ethinyl Estradiol ;
- . . Mestranol / Ethinyl Estradiol 3-Methyl Ether ;
- . . Quinestrol / Ethinyl Estradiol 3-Cyclopentyl Ether ;

Level 2 of 2

Page 0

PRINT FILE ON

SCOPE OFF

ALYZE TI/AB OFF

CHEMTOOL

GLOBAL HELP

SUBJECT BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

Estrogen Replacement Therapy - Current Slot

AGENT

> ■

Images: None

Index drugs.

Do not use default Estrogens unless particularly discussed.

Consider specific Estrogens term.

The following Chemtool terms map to restrictions on this slot:

1-(2-chloroethyl)-1-nitrosocarbonylalanine estradiol-17-ester; clinacterone;

estradiol-17 beta-benzoate; estradiol valerate; KH 2210; Leo 271 f;

stilbestat; **tristequens**;

7(0)-(2,3,5-trideoxy-3-C-formylpentafuranosyl)daun / 7-TFPD ; 7-0-(2,3,5-trideoxy-3-formyl-alpha-L-threo
 N-trifluoroacetyladriamycin / AD 41 ;
 trifluoroacetyladriamycin-14-octanoate / TFAAD ;
 N-trifluoroacetyladriamycinol / AD 92 ; AD-p2 ;
 N-trifluoroacetyladriamycin 14-valerate / AD 32 ; AD-32 ; NSC-246131 ;
 trifluoroacetyl-daunomycin / N-(trifluoroacetyl)daunomycin ;
 trifluoromethylglyoxal-bis(guanyldrazone) / CF3-GBG ;
 2,4,6-trimethylmelamine / N,N',N''-trimethylmelamine ;
 6-trimethylsilylthio-9-trimethylsilyl-purine / S-HP ;
 trimetrexate glucuronate / trimetrexate glucuronic acid combination ;
 tri-1',5',5'-N,N,N-mono-8'-O-methylstreptonigrone / tetrahydrostreptonigrone ; N(1'),N(5'),N(5'),O(8')-tet
 trisrubicinol
 / Kliajest ; Trisequens ;
 trofosamide / ixoten ; N,N,N'-tris(2-chloroethyl)-N'-O-propylene phosphor ; trophosphamide ; Z 4828 ;
 Ts-72 / N-(p-di(2-chloroethyl)aminophenylbutyryl)-p-aminob ; Ts 72 ; Ts72 ;
 tubercidin 3',5'-cyclic phosphate
 tubercidin-5'-diphosphate / TuDP ;
 tubercidin 5'-diphosphate-5'-1,2-dipalmitin / TuDP-dipalmitin ;
 tubercidin 5'-phosphate / tubercidin 5'-monophosphate ; TuMP ;
 tubercidin 5'-triphosphate
 U-30,964 / BENU ; NSC-174793 ; U 30,964 ; U 30964 ; U-30964 ;
 UCH II protocol / UCH-II protocol ;
 UCH I protocol / UCH-I protocol ;
 U 7209SE / 17,21-dihydroxy-11-tert-butylacetoxy-1,4-pregnadiol ; U-7209SE ;
 UFD 1 protocol / UFD-1 ;
 UFTM protocol
 1-UFT protocol / 1 UFT ; 1-UFT ; UFT protocol ;

FILE ON

PE ON

TE TI/AB OFF

FILE

T BY ACTION

CTIONS

RARCHY

PE NOTE

AIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

Estrogen Replacement Therapy - Overview

AGENT

> Trisequens

> ■

NO: triphasic preparation for estrogen replacement therapy; each 28-day
 calendar pack contains 12 tablets of estradiol 2 mg & estriol 1 mg, 10 of
 estradiol 2 mg & estriol 1 mg & norethisterone acetate 1 mg, 6 of estradiol
 1 mg & estriol 0.5 mg

Message Area

Administration - Comment Frame
 ADMINISTRATION/DOSAGE
 nil
 INSTRUMENTATION
 (NIL)
 TOXIC-EFFECT
 nil
 CONTRAINDICATION
 nil
 GENETICS/IMMUNOLOGY
 nil
 PHARMACOLOGY/PHYSIOLOGY
 nil
 BIOSYNTHESIS/METABOLISM/PHARMACOKI
 nil
 >
 HM
 (IEstradiol T)
 (IEstriol T)
 (INorethindrone T
 I/ANALOGS & DERIVATIVES)
 (IDrug Combinations)

Estrogen Replacement Therapy
 METHOD/SUBPROCEDURE
 nil
 INSTRUMENTATION
 (NIL)
 AGENT
 (Irisetate)
 ADVERSE-EFFECT
 nil
 CONTRAINDICATION
 nil
 PROBLEM
 Osteoporosis, Postmenopausal
 Spinal Diseases
 PURPOSE
 (IOsteoporosis, Postmenopausal I/PREVENTION & CONTROL)
 (ISpinal Diseases I/PREVENTION & CONTROL)
 ASSESSMENT
 nil
 STATISTICS/NUMERICAL-DATA

Administration - Comment Slot

PURPOSE
 > THERAPY
 >

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

REVISION

GENERAL HELP

SUBJECT BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

no TIME TERM

Message Area
 Only fillers for this slot:
 /DIAGNOSIS; /THERAPY;

MEDINDEX		20						
Administration/Dosage Frame ADMINISTRATION/DOSAGE nil Therapeutic Index Frame 1)Estrogen Replacement Therapy AGENT (A)		PRINT FILE ON SCOPE OFF ANALYZE TI/AB OFF						
Indices Display PT: MH: *Estrogen Replacement Therapy Human Osteoporosis, Postmenopausal/*PREVENTION & CONTROL Spinal Diseases/ PREVENTION & CONTROL Trisequens/THERAPEUTIC USE Star Rules Save		RESTORE RESTORE FILE RESTORE ACTION						
		RESTRICTIONS HIERARCHY SCOPE NOTE						
		EXPLAIN SLOT						
		INDICES MEDLINE INDICES SAVED MESSAGES CONSIST CHECK						
		SUBJECT FRAME DOCUMENT FRAME						
		CAN/RESTORE FR RESTART DOC CANCEL DOC SEND MAIL ACTIVE TERM						
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%; text-align: center; padding: 2px;">Store Frame</th> <th style="width: 33%; text-align: center; padding: 2px;">Redo Frame</th> <th style="width: 33%; text-align: center; padding: 2px;">Redo Slot</th> </tr> </thead> <tbody> <tr> <td colspan="3" style="height: 120px;"></td> </tr> </tbody> </table>			Store Frame	Redo Frame	Redo Slot			
Store Frame	Redo Frame	Redo Slot						

MEDINDEX		2
Information Frame ADMINISTRATION/DOSAGE nil Information Frame 1) Estrogen Replacement Therapy AGENT [A] Information Frame 1) Estrogen Replacement Therapy AGENT [A]		PRINT FILE ON SCOPE OFF ANALYZE TI/AB OFF
MEDLINE Indices Display PT: MH: Drug Combinations/ THERAPEUTIC USE Estradiol/*THERAPEUTIC USE Estrinol/*THERAPEUTIC USE *Estrogen Replacement Therapy Human Norethindrone/*ANALOGS & DERIVATIVES/ THERAPEUTIC USE Osteoporosis, Postmenopausal/*PREVENTION & CONTROL Spinal Diseases/ PREVENTION & CONTROL		CHENTION GENERAL HELP ABOUT MEDINDEX RESTRICTIONS HIERARCHY SCOPE NOTE EXPLAIN SLOT INDICES MEDLINE INDICES SAVED MESSAGES CONSIST CHECK SUBJECT FRAME DOCUMENT FRAME CAN/RESTORE FR RESTART DOC CANCEL DOC SEND MAIL ACTIVE TERM
Store Frame Redo Frame Redo Slot		

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

1) DRUG THERAPY - CONCEPT FRAME
 METHOD/SUBPROCEDURE
 nil
 INSTRUMENTATION
 (NIL)
 >
 ADVERSE-EFFECT
 nil
 CONTRAINDICATION
 nil
 PROBLEM
 nil
 PURPOSE
 nil
 ASSESSMENT
 nil
 STATISTICS/NUMERICAL-DATA
 nil
 TEMPORALITY
 nil
 PSYCHOLOGICAL/SOCIAL-PROCESS/CONDITION
 nil

1) INITIAL-FRAME INITIAL-MAJOR-TOPIC (A)

RESTRICTIONS
 HIERARCHY
 SCOPE NOTE

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

1) DRUG THERAPY - CONCEPT FRAME

AGENT
 > Trisequens
 > ■

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

Instead of /DRUG THERAPY, consider the following frames:
 Estrogen Replacement Therapy

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

Method/Subprocedure
 nil
Instrumentation
 (NIL)
Agent
 Trisequens
Adverse-Effect
 nil
Contraindication
 nil
Problem
 Osteoporosis, Postmenopausal
 Spinal Diseases
Purpose
 (Osteoporosis, Postmenopausal
 I/PREVENTION & CONTROL)
 (Spinal Diseases
 I/PREVENTION & CONTROL)
 (NIL)
Assessment
 nil
Statistics/Numerical-Data
 nil
Temporality
 nil
Psychological/Social-Process/Condi
 nil

1) INITIAL-FRAME INITIAL-MAJOR-TOPIC (A)

EDITING

GLOBAL EDIT

GLOBAL EDIT

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TEXT

Store Frame

Redo Frame

Redo Slot

Enterogen, Epitaximent, Therapies - Content Frame

METHOD/SUBPROCEDURE

nil

INSTRUMENTATION

(NIL)

AGENT

Trisequens

ADVERSE-EFFECT

nil

CONTRAINDICATION

nil

PROBLEM

Osteoporosis, Postmenopausal

Spinal Diseases

PURPOSE

(!Osteoporosis, Postmenopausal!

! /PREVENTION & CONTROL!)

(!Spinal Diseases!

! /PREVENTION & CONTROL!)

(NIL)

STATISTICS/NUMERICAL-DATA

nil

TEMPORALITY

nil

PSYCHOLOGICAL/SOCIAL-PROCESS/CONDI

nil

Disaggregating Frame

1) INITIAL-FRAME INITIAL-MAJOR-TOPIC (A)

Enterogen, Epitaximent, Therapies - Content Slot

ASSESSMENT

> ■

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

REVISION

RECENT HIER

RECENT ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

Message Area

Standard coords for this slot:

Prognosis, Treatment Failure, Treatment Outcome.

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

Osteoporosis - Concept Frame

BODY-SITE
nil
DISEASE-BY-SITE
nil
AGE-OF-ONSET
nil
>
ETIOLOGY
nil
DISEASE-BY-CAUSATIVE-AGENT
nil
COMPLICATING-PROCESS/CONDITION
nil
BIOLOGICAL-PROCESS/ATTRIBUTE/MEASURE
nil
EPIDEMIOLOGY/DISEASE-STATISTICS
nil
TEMPORALITY
nil
PSYCHOLOGICAL/SOCIAL-PROCESS/CONDITION
nil

Osteoporosis - Concept Frame

1) INITIAL-FRAME INITIAL-MAJOR-TOPIC (A)

RESTRICTIONS

HIERARCHY

SCOPE NOTE

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

Osteoporosis - Concept Frame

PROCEDURE

- > Estrogen Replacement Therapy
- > ■

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERMS

Osteoporosis

Instead of Osteoporosis, consider the following frames:
Osteoporosis, Postmenopausal

Free Radical Scavengers

- . Free Radical Scavengers
- . Growth Substances, Pigments, Vitamins [non-med] (+)
- . Hazardous and Poisonous Substances [non-med] (+)
- . Hematologic, Gastric, Renal Agents [non-med] (+)
- . Heterocyclic Compounds (+)
- . Hormones, Substitutes, Antagonists [non-med] (+)
- . Immunologic and Biologic Factors [non-med] (+)
- . Indicators and Reagents (+)
- . Inorganic Chemicals [non-med] (+)
- . Ion Exchange Resins (+)
- . Irritants / Counterirritants ; Pustulants ; Rubefacients ; Vesicants ;
- . Ligands
- . Lipids and Antilipemic Agents [non-med] (+)
- . Molecular Probes (+)
- . Neurotransmitters and Neurotransmitter Agents [non-med] (+)
- . Nucleic Acids, Nucleotides, and Nucleosides [non-med] (+)
- . **Openic Antibiotics** [non-med] (+)
- . Peripheral Nervous System Agents [non-med] (+)
- . Pharmaceutical Aids (+)
- . Polycyclic Hydrocarbons (+)
- . Radiation-Protective Agents (+)
- . Radiation-Sensitizing Agents (+)
- . Reproductive Control Agents [non-med] (+)
- . Solvents
- . Substances by Action [non-med] (+)
- . Surface-Active Agents (+) / Surfactants ;
- . Tars (+)

Level 1 of 9

Page 1

PRINT FILE ON

SCOPE OFF

ALYZE TI/AB OFF

REPRINT

SCREEN HELP

SUBJECT BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

INITIAL-MAJOR-TOPIC

>

Message Area

Navigation Description

Organic Chemicals (non-med)

- . Alcohols
 - .. Alcohol, Ethyl (+) / Ethanol ;
 - .. Alcohol, Methyl / Methanol ;
 - .. Alcohol, Propyl (+) / Isopropanol ; Isopropyl Alcohol ; Propanol ;
 - .. Alcohols, Amyl / Pentanol ;
 - .. Alcohols, Butyl / Butanols ;
 - .. Alcohols, Hexyl / Hexanols ;
 - .. Alcohols, Octyl / Octanols ;
 - .. Amino Alcohols (+)
 - .. Cyclohexanols (+)
 - .. Fatty Alcohols (+)
 - .. Glycols (+)
 - .. Polyvinyl Alcohol
 - .. Sugar Alcohols (+)
- . Aldehydes
 - .. Formaldehyde / Formalin ;
 - .. Glyoxal
- . Amides
 - .. Acetamides (+)
 - .. Anilides (+)
 - .. Benzanides (+)
 - .. Formanides
 - .. Lactams
 - .. Salicylamides
 - .. Sulfonamides (+)
- . Amidines

Level 2 of 8

Page 1

Return

PRINT FILE ON

SCOPE OFF

ALYZE TI/AB OFF

REINDEX

GLOBAL HELP

OBJECT BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

ACTIVE TERM

INITIAL-MAJOR-TOPIC

>

Select level from pop-up.

Leaving pop-up without selecting is same as selecting DEFAULT 2.

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

INITIAL-FRAME - Current Frame

>
GENDER
Female
AGE
nil
HUMAN/ANIMAL/PREGNANCY/HOMINIDAE
Human
CASE-REPORT/COMPARATIVE-STUDY/IN-V
nil
STUDY-METHOD/DESIGN/FACTOR/MEASURE
nil
PUBLICATION-TYPE
nil
NON-US-GOV/US-GOV-PHS/US-GOV-NON-P
nil

On going study Frame

nil

DEFINITION

GENERAL HIER

QUEST BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INITIAL-FRAME - Current Slot

INITIAL-MAJOR-TOPIC

> Osteoporosis, Postmenopausal
> est rep th
>

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

HIT TIME TERM

Message Area

PRINT FILE ON

SCOPE OFF

ANALYZE TI/AB OFF

INITIAL-FRAME - Current Frame

GENDER

Female

AGE

nil

HUMAN/ANIMAL/PREGNANCY/HOMINIDAE

Human

CASE-REPORT/COMPARATIVE-STUDY/IN-V

nil

STUDY-METHOD/DESIGN/FACTOR/MEASURE

nil

PUBLICATION-TYPE

nil

NON-US-GOV/US-GOV-PHS/US-GOV-NON-P

nil

Outstanding Frame

nil

Estradiol

Estradiol Antagonists

Estranes

Estrenes

Estriol

Estrogen Antagonists

Estrogen Replacement Therapy

Estrogens

Estrogens, Catechol

Estrogens, Conjugated

Estrogens, Non-Steroidal

Estrogens, Synthetic

Estrone

Ethinyl Estradiol

Receptors, Estrogen

estradiol-17 beta-benzoate

estradiol mustard

estradiol valerate

Leo 271 f

INITIAL

> Ost

> estr

>

REVISION

REVIEW HELP

QUEST BY ACTION

RESTRICTIONS

HIERARCHY

SCOPE NOTE

EXPLAIN SLOT

INDICES

MEDLINE INDICES

SAVED MESSAGES

CONSIST CHECK

SUBJECT FRAME

DOCUMENT FRAME

CAN/RESTORE FR

RESTART DOC

CANCEL DOC

SEND MAIL

Reference Terms

Estradiol / 17 beta-Estradiol ; Estradiol-17 beta

Estradiol Antagonists

Estranes / 19-Norandrostanes

Estrenes / 19-Norandrostenes

Estriol

Estrogen Antagonists / Antiestrogens

Estrogen Replacement Therapy / Postmenopausal Hormone Replacement Therapy

Estrogens

Estrogens, Catechol

Estrogens, Conjugated / Premarin

Estrogens, Non-Steroidal

Estrogens, Synthetic / Estrogen Analogs

Estrone

Ethinyl Estradiol / Ethinyl Estradiol

Receptors, Estrogen

Message Area

Postmenopausal - Current Frame

Original Frame

BODY-SITE

Spine

DISEASE BY SITE

1) INITIAL-FRAME INITIAL-MAJOR-TOPIC (A)

2) Estrogen Replacement Therapy PROBLEM (R)

3) Absorptionmetry, Photon PROBLEM (R)

Indices Display

JOURNAL ARTICLE

Absorptionmetry, Photon

Administration, Oral

Bone Density/ DRUG EFFECTS

Densitometry, X-Ray

Drug Administration Schedule

Drug Implants

*Estrogen Replacement Therapy

Estrogens/ ADMINISTRATION & DOSAGE/ PHARMACOLOGY

Female

Fractures/ ETIOLOGY/ PREVENTION & CONTROL

Human

Injections, Intravenous

Menopause

Middle Age

Monitoring, Physiologic

Osteoporosis, Postmenopausal/ COMPLICATIONS/*PREVENTION & CONTROL/ RADIOGRAPHY/

Spinal Diseases/ PREVENTION & CONTROL/ RADIOGRAPHY/ RADIONUCLIDE IMAGING

Tomography, X-Ray Computed

Done

Ignore

Message Area

PRINT FILE ON

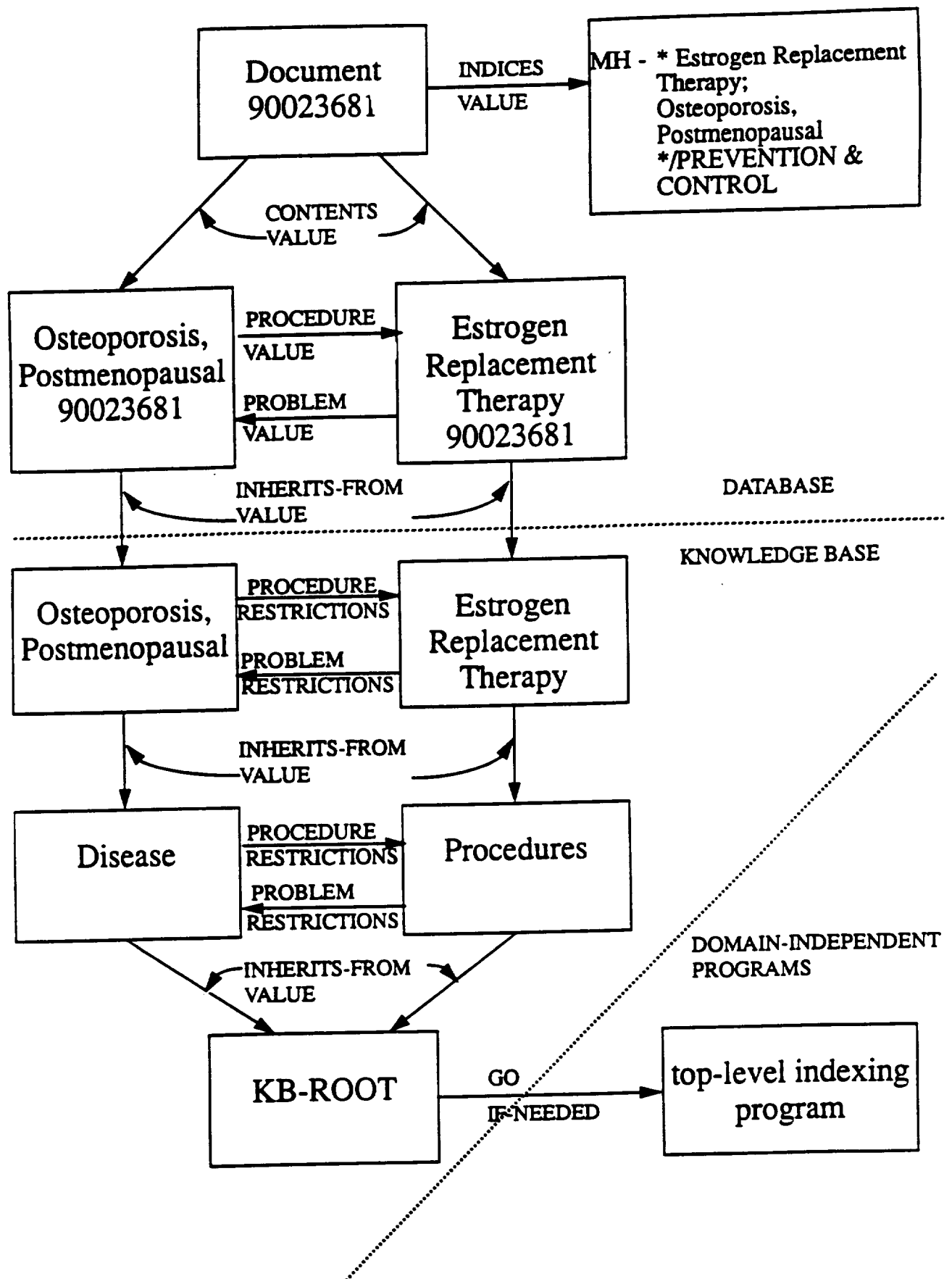
"Monitoring skeletal response to estrogen."

"Estrogen replacement therapy at accepted doses is not fully effective in preventing bone loss and fractures in postmenopausal women. Bone densitometry is useful for monitoring estrogen replacement therapy to assess dose, foster compliance and check for secondary bone loss. The most appropriate site for bone loss monitoring is probably the spine because it shows larger decreases at the menopause than appendicular sites, it shows larger increases with therapy, and it has clinical import in terms of fracture. Both dual-photon absorptiometry (or dual-energy x-ray absorptiometry) and computed tomography are the preferred monitoring methods. The precision of these densitometry methods is generally adequate to permit interim decisions with regard to continuing therapy, as well as conclusive decisions on therapeutic efficacy after 1 to 2 years of monitoring. Judicious use of densitometry in combination with biochemical determinations can enhance therapeutic control and provide both patient and physician confidence in long-term estrogen replacement therapy."

("SKELETAL" (NIL))
("ESTROGEN REPLACEMENT THERAPY" (Estrogen Replacement Therapy))
("EFFECTIVE" (NIL))
("FRACTURES" (Fractures))
("DOSE" NIL)
("SECONDARY" NIL)
("BONE LOSS" NIL)
("SPINE" (Spine))
("MENOPAUSE" (Menopause))
("CLINICAL" NIL)
("DUAL PHOTON ABSORPTIOMETRY" (Absorptiometry, Photon))
("DUAL ENERGY X-RAY ABSORPTIOMETRY" (Densitometry, X-Ray))
("COMPUTED TOMOGRAPHY" NIL)
("PREFERRED" (NIL))
("METHODS" (METHODS))
("CONTINUING" (NIL))
("WELL" (NIL))
("2 YEARS" (Child, Preschool))
("MONITORING" NIL)
("COMBINATION" NIL)
("BIOCHEMICAL" NIL)
("CONTROL" (PREVENTION & CONTROL))
("PATIENT" (Patients))
("PHYSICIAN" (NIL))
("LONG TERM" (NIL))

INITIAL-FRAME - Current Slot	
INITIAL-MAJOR-TOPIC	
>	Densitometry, X-Ray
>	Absorptiometry, Photon
>	Menopause
>	Fractures
>	Estrogen Replacement Therapy
>	■

Automated indexing of sample title and abstract based on word matching against knowledge base terms and aliases.



Indexing in MedIndEx is the process of:

Filling indexing frames which inherit from KB frames

New indexing frame has only two slots with values:

Estrogen Replacement Therapy 90136159

INHERITS-FROM:

VALUE: Estrogen Replacement Therapy

UNIQUE-ID:

VALUE: 90136159

Estrogen Replacement Therapy 90023681
inherits slots (with its data and programs) from:

Estrogen Replacement Therapy

INHERITS-FROM:

VALUE: Drug Therapy

PROBLEM: <no VALUE>

RESTRICTIONS:

**<program: VALUE must be a member
of DISEASE-LIST>**

DEFAULT: Osteoporosis, Postmenopausal

*** IF-NEEDED:**

**<program: if there is a DEFAULT,
display it as a VALUE>**

CHILDREN: <no VALUE>

*** inherited from an ancestral frame in the
inheritance hierarchy**

Specificity Encoded in Frames

Fractures 90136159 (indexing frame)

**INHERITS-FROM value: Fractures
(established by system)**

BODY-SITE value: Spine (added by indexer)

**When indexer adds Spine to BODY-SITE value in
Fractures indexing frame, system says Spinal
Fractures must be used instead.**

Fractures (KB frame)

**CHILDREN specificity:
BODY-SITE Spinal Fractures**

Spinal Fractures (KB frame)

BODY-SITE specificity: Spine

MedIndEx (Medical Indexing Expert) frames:

Knowledge Base (KB) frames

- **Encode concept/rules**
- **Linked by domain-specific relations to provide knowledge-based and situation-specific assistance**
- **Inheritance classification**
- **Conventional classification (“trees”)**

Indexing frames

- **Inherit procedures / data from KB frames**
- **Contain the data for generating MEDLINE indexing**

Presumed Advantages of Knowledge-Based Indexing

**Rules as system-initiated executable
procedures**

Inheritance for less redundancy, more rules

Improved consistency in applying rules

**Document-specific information used for
situation-specific assistance**

Domain relationships used by rules

Additional Significant Points about MedIndEx

Self-training/teaching

- **Reinforces learning**
- **Conveys rule changes quickly and directly**

Electronic mail originating within system

- **Serves as electronic suggestion box**
- **Can include copies of frames**

Foundation for managing and expanding the indexing scheme

- **Addresses gaps, inconsistencies, vagueness**
- **Facilitates quickly trying out indexing concepts**

Significant Points (continued)

Integration of applications in one system

- **Indexing**
- **Searching**

Integration of additional knowledge sources into the system

Portability across:

- **Subject domains**
- **Type of material indexed (e.g., images)**

Machine learning from database facilitated by precision of links

- **Updating KB classification scheme**
- **Offering users intermediate selection based on probability**

Standardized, readily-available software (X Windows, Common Lisp, Unix)

MedIndEx “Processes”

- **Journal Assignment Interface**
- **Indexer/Searcher Interface**
- **Indexing Score Report Generator**
- **KB Manager Interface**
- **KB Manager Tools**

KB Manager Purpose and Function

Purpose:

- **Ensure proper syntax**
- **Ensure consistency**

Function:

- **Create new KB frames**
- **Make inheritance links between KB frames**
- **Make children links (similar to *MeSH Trees*) between KB frames**
- **Encode indexing assistance in KB frames**

KB Manager Batch Mode Input

MeSH Tree Structures

000 Disease

000.C14 Cardiovascular Diseases

000.C14.280 Heart Diseases

000.C14.280.459 Heart Neoplasms

000.C14.907 Vascular Diseases

000.C04 Neoplasms

000.C04.588 Neoplasms by Site

000.C04.588.894 Thoracic Neoplasms

000.C04.588.894.309 Heart Neoplasms

KB Manager Batch Mode Input (continued)

Art and Architecture Thesaurus

Single Built Works and Open Spaces

- ..<single built works>**
- <single built works by specific type>**
- <single built works by function>**
- <religious structures>**
- religious buildings**
- churches**
- <churches by location or context>**
- cathedrals**

MEDLINE Search Assistant

- **Uses query frames from MedIndEx**
- **Generates MEDLINE indexing terms from query frames**
- **Generates terms as suggested MEDLINE search terms**
- **Provides interface to develop search strategy**
- **Uses search syntax of public retrieval system**
- **Connects to retrieval system and runs search**

Sample Query

Monitoring bone density changes in postmenopausal osteoporosis treated by estrogen replacement therapy.

((explode Estrogens/THERAPEUTIC USE **OR** explode Estrogen Replacement Therapy
OR explode Estrogens/CONTRAINDICATIONS
OR explode Estrogens/ADVERSE EFFECTS
OR explode Estrogens/ADMINISTRATION & DOSAGE
OR explode Estrogens/POISONING))

AND

(explode Osteoporosis, Postmenopausal **OR** explode Menopause)

AND

(explode Bone Density **OR** explode Densitometry))

Evaluation

Funding

NIH 1% Evaluation Set-Aside Fund,
Section 513, Public Health Service Act

Evaluation Design Project (completed)

Development of Experimental Design to
Establish a Methodology for Evaluating
Knowledge-Based Systems for Subject
Indexing. NIH/NLM 92-308

“Design for a Study to Evaluate the
MedIndEx Approach to Subject Indexing.”
Prepared by Herner and Company,
Decision Sciences Associates, and the
University of Illinois Graduate School of
Library and Information Science. NTIS
PB93-236016.

Evaluation Project (ongoing)

Evaluation of a Knowledge-Based Expert
System for Subject Indexing. NIH/NLM
94-305.

Evaluation

Objectives

Primarily, to determine if MedIndEx is sufficiently promising as an approach to MEDLINE indexing to warrant further development

Secondarily, to identify ways in which MedIndEx can be improved

Evaluation

Parameters

Comparison of MedIndEx and AIMS (indexing environment variable)

indexing quality

inter-indexer consistency

Effects of other variables on quality and consistency

task knowledge (MEDLINE indexing)

tool knowledge (gained through use)

User reaction to MedIndEx

Average time needed to index a document

Time-stamped transaction logs

Within test domain, some diversity in subject matter and indexing complexity of test documents

Evaluation

Components

Indexing experiment to compare:

MedIndEx output (quality and consistency) for different types of indexers (experienced and inexperienced) and for different levels of experience with the system

MedIndEx output with AIMS output

A standard for quality of indexing and a method for judging quality of testing indexing against this standard

Method of scoring inter-indexer consistency

Criteria and methods for identifying set of test documents

Survey instrument to determine user reactions to MedIndEx

Procedures for capturing data for logs

Method for recording time to index a document

Evaluation

Test Design

Classic split-plot factorial design (Kirk, 1982)

Four groups characterized by task knowledge (low, high) and system used (MedIndEx, AIMS)

Both MedIndEx groups and low task knowledge AIMS group index three sets of ten documents; high task knowledge AIMS group indexes only third set

Cell size:

low task knowledge groups - 15

high task knowledge groups - 10

Evaluation

Test Design

15 hypotheses specifying groups compared

four hypotheses each for quality, consistency, and time; three hypotheses for attitudinal differences

Examples of hypothesis for quality and consistency:

H1: Indexing produced using MedIndEx is significantly different in terms of quality from indexing produced using AIMS

H5: Indexing produced using MedIndEx is significantly different in terms of consistency from indexing produced using AIMS

H1 and H5 groups: low task knowledge MedIndEx versus low task knowledge AIMS, high task knowledge MedIndEx versus high task knowledge AIMS

Evaluation

Test Design

Training schedule

inexperienced groups - five days indexing
training in test domain

inexperienced MedIndEx and AIMS groups -
four days hands-on training, six days indexing
three sets of ten test documents, questionnaire
after sets 1 and 3

experienced AIMS group - half-day indexing
document set 1

Evaluation

Criteria for Evaluating Indexing Quality

Positive scoring (type of term, starred vs unstarred, number of points added):

main headings starred; no starred subheading -
20 if starred, 6 if unstarred, 0 if absent

main headings starred; at least one starred
subheading - 15 if starred, 6 if unstarred, 0 if
absent

subheadings starred - 8 if starred, 5 if
unstarred, 0 if absent

check tags - 7, always unstarred, 0 if absent

main headings unstarred - 6 if unstarred, 5 if
starred, 0 if absent

subheadings unstarred - 5 if unstarred, 4 if
starred, 0 if absent

Evaluation

Criteria for Evaluating Indexing Quality

Negative scoring (type of term, starred vs unstarred, number of points deducted):

main headings starred - 7

subheadings starred - 5

check tags (always unstarred) - 4

main headings unstarred - 1

subheadings unstarred - 1

Evaluation

Criteria for Evaluating Indexing Quality

Example of standard and four test indexers:

Standard	Indexer A	Indexer B
MH1*	MH1*	MH1*
MH2	MH2	
25	$25/25=1.00$	$20/25=.80$

Indexer C	Indexer C
MH1*	MH1*
MH3	MH3*
$19/25=.76$	$13/25=.52$

Evaluation

Criteria for Evaluating Inter-Indexer Consistency

Consistency of a Pair (CP) of indexers =

number of term agreements

number of unique terms assigned

Consistency of a Group (CG) is mean of all pairwise scores (CP's for each pair)

Rodgers, 1961, and Hooper, 1965

Evaluation

Questionnaire

68 short answer questions

34 attributes (Adelman, Ulvila, and Lehner, 1991)

attribute categories: judgmental performance, observable use, opinion on usability, scope of application, explanation, and organizational impact

11-point Likert scale from 0 (very strongly disagree) to 10 (very strongly agree)

Each attribute represented in positive and negative statements

High scores on positive statements and low scores on negative statements indicate good performance on the attribute

four open ended questions

MedIndEx Requirements

Sun SPARCstation 2

64 Mbytes memory, Gbyte disk

SunOS operating system

Lisp files

Lucid Common Lisp 4.1

CLX and CLUE (public domain from TI)

MedIndEx files

X11 Release 5

twm window manager

(experimental) PC, Windows 3.1, and modem

eXceed/Xpress-W (X Windows server)

graphics card for 1280x1024 resolution

**Indexing, Providing Access to Information:
Looking Back, Looking Ahead,
The Proceedings of the 25th Annual Meeting
of the American Society of Indexers**

Nancy C. Mulvany, editor

**American Society of Indexers, Inc.
Port Aransas, Texas
1993**

Cover Art © 1992 André Kuzniarek

Copyright © 1993 by the American Society of Indexers, Inc. All rights reserved. Printed in the United States of America. No part of this book may be reproduced in any form by any electronic or mechanical means including information storage and retrieval systems without express written permission from the society, except by a reviewer who may quote brief passages for a review.

**Published by
The American Society of Indexers, Inc.
P.O. Box 386
Port Aransas, TX 78373**

ISBN 0-936547-19-7

The MedIndEx™ Prototype for Computer Assisted MEDLINE® Database Indexing

Susanne M. Humphrey

National Library of Medicine, Bethesda, MD 20894, USA

Abstract

The MedIndEx (Medical Indexing Expert) research project combines artificial intelligence and information retrieval principles and methods to develop and test an interactive knowledge-based prototype for computer-assisted indexing of the MEDLINE database. By encoding the indexing scheme and the MeSH® (Medical Subject Headings) thesaurus in a knowledge base (KB), and designing a system for indexers to use in a workstation environment, the objective of this project is to facilitate "expert indexing" that is performed at the National Library of Medicine (NLM).

Introduction

MedIndEx is a prototype knowledge-based expert system designed to assist indexers in creating subject access points for performing MeSH thesaurus-based searches of the MEDLINE database [1]. MEDLINE contains more than 6.6 million citations covering the periodical biomedical literature since 1966. In 1992, NLM indexed more than 400,000 documents for MEDLINE. Over two million on-line searches were performed on the current MEDLINE database (the most recent 2-3 years) using NLM's retrieval system.

The main objective of MedIndEx is to develop interactive knowledge-based systems to facilitate expert indexing that goes into the MEDLINE product. The system consists of computer representations of MeSH concepts, enhanced by domain-specific relations; executable rules based on NLM's indexing scheme; and an interface designed for a sophisticated workstation environment.

The aim of this paper is to provide a sense of the potential for the knowledge-based approach to database indexing. The next section describes conventional database indexing as background. This is followed by the section describing MedIndEx. The concluding section is a brief statement of the current status of the project.

Conventional Indexing

This section summarizes, describes, and illustrates features of conventional indexing that carry over to a knowledge-based approach as a natural outgrowth. These features are:

- Basic tenets of database indexing
- Thesaurus, including explicit classification and implied relations
- Indexing rules
- Windowing/mousing environment

Indexing tenets defined in NLM's indexing manual include specificity and multiplicity. These correspond to the following fundamental rules of indexing identified by Lancaster [2]:

- Include all the topics known to be of interest to the users of the information service that are treated substantively in the document
- Index each of these as specifically as the vocabulary of the system allows and the needs or interests of the users warrant

The first rule corresponds to the conceptual analysis stage of indexing; the second, to the translation stage [3].

The MeSH thesaurus contains about 17,000 headings and 80 subheadings which may be appended to headings to form heading-subheading pre-coordinations as indexing terms. An example of such a pre-coordination is Estrogens /THERAPEUTIC USE. In addition, MeSH contains official aliases (synonyms, abbreviations, lexical variants, etc.)

Indexing rules are documented in various published tools, including the indexing manual, technical notes, and the alphabetic MeSH in the form of annotations appended to individual terms. NLM uses a system of *coordinate indexing*, and virtually all the rules reflect this orientation. Coordination rules may be illustrated by the example of indexing the following topic:

estradiol in estrogen replacement therapy for prevention of spinal bone loss in postmenopausal women

This would be indexed by the following indexing terms:

Estrogen Replacement Therapy

Estradiol /THERAPEUTIC USE

Osteoporosis, Postmenopausal /PREVENTION & CONTROL

Spinal Diseases /PREVENTION & CONTROL

Two terms refer to a disease, Osteoporosis, Postmenopausal and Spinal Diseases; one term, Estrogen Replacement Therapy, to a medical procedure; and the remaining term, Estradiol, to a drug. Taken together, these four terms describe the sample topic, and serve to illustrate coordination of terms from different MeSH categories (Diseases, Techniques and Equipment, Chemicals and Drugs) to index important ideas in a document. Insofar as this coordination may express the general topic of the document, it also illustrates the multiplicity tenet of indexing, mentioned earlier.

Coordination is also used for expressing the individual topics in the phrase. This can be done by using a pre-coordinated heading that expresses two or more concepts in the domain being indexed. In this example, Estrogen Replacement Therapy, which expresses estrogen as a therapeutic agent for conditions requiring its replacement (estrogen deficiency), is the most complex pre-coordinated heading in this example, in terms of the number of categories represented in this term. Another form of coordination is the heading/subheading pre-coordination, of which there are three in this example.

Finally, coordination to index specific topics is achieved by combining different headings, as it is for expressing the general topic. Sometimes these sorts of coordination use special indexing rules. Furthermore, elements of the coordination may themselves be pre-coordinated. In this case, one might surmise that *spinal bone loss* would be expressed by combining Osteoporosis (the medical term for this condition) and Spine. However, MEDLINE indexing rules dictate that the organ-disease pre-coordinate is a more precise expression of this topic than the organ, hence, Spinal Diseases is used rather than Spine. And, because of the qualification *in postmenopausal women*, the Osteoporosis, Postmenopausal pre-coordinate is preferred over Osteoporosis.

Features of conventional indexing contain implied relations that can be quite useful in a knowledge-based approach. For example, imagine that you are indexing in the area of drug therapy. The MeSH categories would be useful as reminders for indexing the various aspects of this area. The Diseases category would be used for covering the condition being treated; the Chemicals and Drugs category, for the drug being used. Imagine now that you are indexing in the area of drugs. The MeSH topical subheadings would also be useful as prompts for aspects to be covered. For example "administration & dosage" would remind you of terms like Injections, Drug Administration Schedule, and Drug Therapy, Combination; "adverse effects" would alert you to covering the drug's toxicity.

Coordination rules in the indexing documentation suggest relations that might be of general use. For example, the MeSH annotation to index *pelvic lymph nodes* by coordinating Lymph Nodes and Pelvis implies a "location" relation. This would be useful as a prompt when a term like Lymph Nodes is assigned, as a reminder to coordinate with a locational term. An important rule for MEDLINE indexing is the "neoplasm coordination rule" which states that to properly index cancer, an organ-neoplasm term should be coordinated with a histologic-type term. For example, *adenocarcinoma of bone*, should be indexed Bone Neoplasms and Adenocarcinoma. Stated in terms of the MeSH classification, this rule suggests the coordination of terms from two major nodes in the Neoplasms category, namely, Neoplasms by Site (for Bone Neoplasms) and Neoplasms by Histologic Type (for Adenocarcinoma). The implied relation "histologic-type" would be useful as a prompt when an organ-neoplasm term is assigned and "neoplasm-by-site" might serve as a prompt for an organ-neoplasm term when a histologic type term is assigned.

Finally, conventional indexing is performed in an interactive environment using NLM's Automated Indexing and Management System (AIMS), introduced in 1984. This system provides important functions previously performed in batch mode, like validating terms and substituting preferred forms for official aliases that may have been entered. It also helps with check tags; for example, the system automatically adds the term Female when the indexer has added Pregnancy, and then displays a message asking the indexer to add Human or Animal. Indexers may request displays of information from MeSH, such as the scope note, annotation, or permissible subheadings for an individual term. The system can be programmed to display help messages for specified terms. For instance, when an array of specific Mastectomy terms was added to MeSH, the system began issuing a message alerting indexers to the availability of these terms whenever the term Mastectomy was assigned. In 1991, NLM began furnishing staff with PS/2® microcomputers, providing an environment with multiple simultaneous windows and mousing. As a result, indexers could simultaneously run the AIMS window and in another window run NLM's retrieval system to search MeSH and MEDLINE, and could cut and paste between windows. Plans are under way to install an OS/2®-compatible version of the WordPerfect® text editor with the indexing manual, technical notes, and other indexing tools on a server. The greatest significance of these developments for MedIndex is that they pave the way for knowledge-based indexing, which requires sophisticated interface technology.

As described in this section, conventional systems inherently use potentially encodable procedural knowledge, illustrated by various coordination rules. Current expression of this knowledge uses existing factual knowledge, in particular the MeSH classification scheme, and relations that are at least implicit throughout the system. Potentially usable relations may be found as implied relations between MeSH categories, in subheadings that might be used for linking terms to one another, and in single pre-coordinated terms. Thus, the knowledge inherent in conventional systems, coupled with installation of workstation-like environments for indexers, provides an important foundation for developing knowledge-based indexing systems.

MedIndex Indexing

Knowledge-based systems for indexing require a new type of data structure that can, in effect, encode the combined factual and procedural knowledge established in the conventional system. The factual knowledge needed for MEDLINE indexing is encoded in MeSH records. But there is very little procedural knowledge encoded in the indexing system. Most indexing rules are provided as textual information. While interfaces may be developed to display this information, a knowledge-based system characteristically does more than this. It uses relations in the domain to provide interactive, situation-specific assistance. As with published cross-references in thesauri, this assistance may be prescriptive or merely suggestive[4, 5]. The difference is that assistance is prompted by users' interactions with the system. The computer "knows" what the user is doing and instructs the user substantively and specifically in continuing the task at hand.

In MedIndEx, a data structure known as a *frame* is used for providing this knowledge-based assistance. A knowledge base frame serves as a template for filling out an indexing form. Rather than explain frames in a technical manner, we will illustrate their use in MedIndEx by example. Figure 1 shows a computer screen containing an indexing frame titled Estrogen Replacement Therapy. This term has been entered in a previous frame as the initial topic of the document being indexed, as shown in the Display window. The current frame name is in the title bar of two of the windows, the Current Frame window and the Current Slot window. The terms in upper case in the Current Frame window are subdivisions of frames, known as *slots* (AGENT, PURPOSE, ASSESSMENT, etc.). These serve as prompts for further information in relation to the frame term (name of the frame).

MEDINDEX SCOPE OFF

CURRENT FRAME

AGENT
nil

PURPOSE
nil

ASSESSMENT
nil

METHOD/SUBPROCEDURE
nil

INSTRUMENTATION
nil

ADVERSE-EFFECT
nil

CONTRAINDICATION
nil

TEMPORALITY
nil

INITIAL-FRAME

INITIAL-MAJOR-TOPIC
Estrogen Replacement Therapy

SEX
Female

AGE
Middle Age

ALPHABETIC/PREGNANCY/MONITORING
Human

CASE-REPORT/COMPARATIVE-STUDY/IN-VITRO
nil

PUBLICATION-TYPE
JOURNAL ARTICLE
CLINICAL TRIAL, PHASE III
RANDOMIZED CONTROLLED TRIAL
NON-IG-GOV/IG-GOV-PHS/IG-GOV-NON-PHS
nil

PROBLEM
> ☐

PRINT FILE ON

Figure 1. Basic Estrogen Replacement Therapy indexing frame.

Estrogen Replacement Therapy

INHERITS-FROM

VALUE: Drug Therapy

PROBLEM

VALUE: None (to be supplied in indexing frame)

RESTRICTIONS:

<program: **VALUE** must be a member of **PROBLEM-LIST**>

DEFAULT: Osteoporosis, Postmenopausal

*** IF-NEEDED:**

<program: if there is a **DEFAULT**, make it available for display>

CAN-ADD?:

<program: if **VALUE** entry is Osteoporosis, create the message "Osteoporosis, Postmenopausal is available.">

AGENT

VALUE: None (to be supplied in indexing frame)

DEFAULT: Estrogens

**** IF-NEEDED:**

<program: if there is a **DEFAULT**, create a warning message not to use it unless particularly discussed, and to consider a specific of it, e.g., "Do not use Estrogens unless particularly discussed. Consider specific Estrogens term.">

PURPOSE

VALUE: None (to be supplied in indexing frame)

**** IF-ADDED:**

<program: update the MEDLINE indexing by appending /THERAPEUTIC USE to the **VALUE** of **AGENT**, and appending /PREVENTION & CONTROL to the **VALUE** of **PROBLEM**>

*** inherited from Procedures frame, top-level frame in KB hierarchy**

**** inherited from Administration of Agent frame, higher-level frame in KB hierarchy**

Figure 2. Depiction of Estrogen Replacement Therapy knowledge base Frame, serving as template for all Estrogen Replacement Therapy indexing frames.

All input from the user is directed by the system into the Current Slot window. The PROBLEM prompt in this window is waiting for the user to enter an indexing term for the medical problem in relation to Estrogen Replacement Therapy. That is, a frame prompts for indexing terms, known as *fillers*, in relation to the frame term (name of the frame), according to the current slot. When the indexer has filled in the slot with appropriate indexing terms, according to the content of the document, the slot with its fillers is returned to the Current Frame window, and the next slot is presented in the Current Slot window for filling.

Figure 2 shows a simplified knowledge base frame for Estrogen Replacement Therapy. Remember, the indexing frame in Figure 1 uses this knowledge base frame as a template. Note the subdivisions of the PROBLEM slot. The first subdivision, labeled RESTRICTIONS, contains a computer program that includes a list of all of the indexing terms that are legal for this slot (primarily from the Diseases category). Furthermore, this program checks that any indexing term a user might enter as a filler for this slot in an indexing frame using this template is a member of this list of permissible terms. The restrictions list of terms is encoded in a form that translates into a hierarchical display. Figure 3 is again the same indexing frame where this display is shown in a Restrictions Display window.

The relationship between an indexing frame and its corresponding knowledge base frame introduces the notion of *inheritance*. As shown in Figure 4, an indexing frame *per se* initially contains only two slots, INHERITS-FROM and UNIQUE-ID. The latter has as its filler an accession number for the document being indexed, and is not significant in this discussion. On the other hand, the INHERITS-FROM slot, having a knowledge base frame term as a filler, links the indexing frame to this knowledge base frame. This link enables the indexing frame to access all the slots and their rules and data from the corresponding knowledge base frame. Also, as seen in Figure 4, internally the name of an indexing frame includes the accession number of the document being indexed. All the data seen on the screen in the Estrogen Replacement Therapy indexing frame (Figure 1) are fetched from the Estrogen Replacement Therapy knowledge base frame by virtue of this INHERITS-FROM link. Only the slots to which the indexer has added a filler in completing an indexing frame (PROBLEM, PURPOSE, and AGENT slots in Figure 5) are physically part of the indexing frame. All the indexing assistance provided by the system is encoded in, and accessed from, knowledge base frames.

Subdivisions of domain-specific slots in Figure 2 will now be explained in terms of providing system-initiated help. A slot in a knowledge base frame may have a DEFAULT filler. Let us say that Estrogen Replacement Therapy is used, in the vast majority of case in the real-world (possibly as determined by examining the MEDLINE database), for treating postmenopausal osteoporosis. By making Osteoporosis, Postmenopausal a DEFAULT in the PROBLEM slot, the system uses a program in the IF-NEEDED subdivision for automatically displaying this default term as a filler for this slot in the indexing frame, and all the indexer need do is verify that it is appropriate for the current document. But suppose this is not an appropriate default, and there are various conditions treated by estrogen replacement therapy, and the user would end up erasing this filler much of the time. Another possible subdivision for the PROBLEM slot is CAN-ADD? This subdivision contains a program with the following rule: If the user enters the filler *Osteoporosis* at the PROBLEM prompt in an Estrogen Replacement Therapy indexing frame, the system generates the following message for display: "Osteoporosis, Postmenopausal if available." Another DEFAULT is Estrogens in the AGENT slot. This time the IF-NEEDED creates a warning message not to use this term unless particularly discussed and to consider specific Estrogens (whereupon the user may request a hierarchical display as was done for the PROBLEM slot earlier).

The data and programs associated with IF-NEEDED, DEFAULT, and CAN-ADD? just described provide suggestive knowledge-based assistance to the indexer in support of the specificity tenet of indexing. The multiplicity tenet is supported by the very nature of a system using slot prompts serving to remind the indexer of aspects of a topic which are consistent with the indexing scheme.

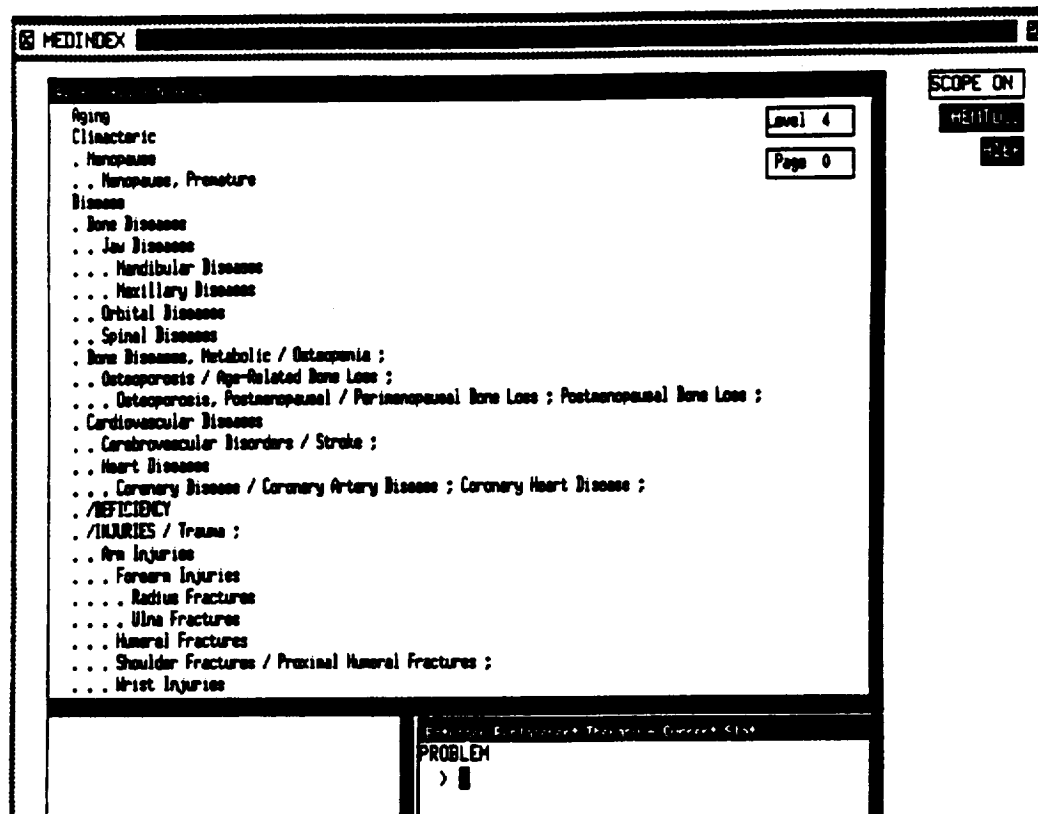


Figure 3. Restrictions display for PROBLEM slot in Estrogen Replacement Therapy indexing frame.

Estrogen Replacement Therapy 90136159

INHERITS-FROM

VALUE: Estrogen Replacement Therapy

UNIQUE-ID

VALUE: 90136159

Figure 4. Depiction of internal Estrogen Replacement Therapy indexing frame before any fillers have been added, showing the inherits-from link to a knowledge base frame.

MEDINDEX	
PROBLEM Osteoporosis, Postmenopausal Spinal Diseases AGENT (Spinal Diseases Estradiol) (Osteoporosis, Postmenopausal Estradiol) (NIL) ASSESSMENT nil METHOD/SUBPROCEDURE nil INSTRUMENTATION nil ADVERSE-EFFECT nil CONTRAINDICATION nil TEMPORALITY nil	INITIAL-FRAME INITIAL-MAJOR-TOPIC Estrogen Replacement Therapy GENDER Female AGE Middle Age HUMAN/NINHAL/PREGNANCY/NONININE Human CASE-REPORT/COMPARATIVE-STUDY/IN-VITRO nil PUBLICATION-TYPE JOURNAL ARTICLE CLINICAL TRIAL, PHASE III RANDOMIZED CONTROLLED TRIAL NON-US-GOV/US-GOV-PHS/US-GOV-NON-PHS nil Enter PURPOSE in relation to: Estradiol for Spinal Diseases > /PREVENTION & CONTROL > Estradiol for Osteoporosis, Postmenopausal > /PREVENTION & CONTROL >

Figure 5. Estrogen Replacement Therapy indexing frame with fillers resulting in system-generated MEDLINE indexing in Figure 6.

MEDINDEX	
ANALOG/DERIVATIVE-P nil ADMINISTRATION-DOSAGE	Estrogen Replacement Therapy
Indexing Display PT: CLINICAL TRIAL, PHASE III JOURNAL ARTICLE RANDOMIZED CONTROLLED TRIAL MH: Administration, Cutaneous Administration, Oral Comparative Study Estradiol/ ADMINISTRATION & DOSAGE/*THERAPEUTIC USE *Estrogen Replacement Therapy Female Human Middle Age Osteoporosis, Postmenopausal/*PREVENTION & CONTROL Spinal Diseases/ PREVENTION & CONTROL	

Figure 6. System-generated MEDLINE indexing based on filled indexing frames, including the frame in Figure 5.

Getting back to inheritance, it would be most inefficient to encode certain rules in all knowledge base frames where they might apply. Estrogen Replacement Therapy is one of numerous frames for medical procedures with which it shares many indexing rules. Furthermore, what is the use of these indexing frames, since there is no retrieval language available for searching them. Both these concerns are addressed by examining another slot subdivision known as IF-ADDED, and showing how it is used.

The most practical output of MedIndEx is conventional MEDLINE indexing. This indexing is generated automatically by the system, based on fillers that have been entered by the user in indexing frames, and is updated as the user proceeds in filling frames. Assuming several indexing frames have been completed, we see in Figure 6 the system-generated indexing based on these frames. To illustrate how this was done, we focus on the terms beginning with Estradiol and Osteoporosis, Postmenopausal, in particular the subheadings THERAPEUTIC USE appended to the former, and PREVENTION & CONTROL to the latter. This indexing was generated based on the Estrogen Replacement Therapy indexing frame in Figure 5, specifically the PURPOSE slot. The rule is encoded as a computer program in the IF-ADDED subdivision of the PURPOSE slot, as shown in Figure 2. Based on IF-ADDED rules associated with earlier slots in the frame, Osteoporosis, Postmenopausal (the PROBLEM filler) and Estradiol (the AGENT filler) have already been added to the conventional MEDLINE indexing. The current IF-ADDED rule (in PURPOSE), based on the filler /PREVENTION & CONTROL, causes /THERAPEUTIC USE to be appended to Estradiol, and /PREVENTION & CONTROL to be appended to Osteoporosis, Postmenopausal.

Not only do indexing frames inherit from knowledge base frames, but knowledge base frames inherit from one another, along an inheritance hierarchy. Since this IF-ADDED rule has wider applicability than just this current frame, it is encoded in a frame as high as possible in the inheritance hierarchy, namely, the Administration of Agent frame, which is three levels above the Estrogen Replacement Therapy frame. Specifically, this rule is inherited via the following chain: Estrogen Replacement Therapy 90136159 INHERITS-FROM Estrogen Replacement Therapy INHERITS-FROM Drug Therapy INHERITS-FROM Drug Administration INHERITS-FROM Administration of Agent. Not only does inheritance save space, but it also helps to ensure consistency by eliminating the need to maintain multiple copies of rules. Furthermore, unless over-ridden, this rule will automatically be inherited by any new KB frame as soon as it joins the Administration of Agent hierarchy.

System assistance described earlier, suggesting specific terms and offering advice about defaults, is also inherited. As seen in Figure 2, the program associated with the IF-NEEDED subdivision of the AGENT slot is also inherited from the Administration of Agent frame. That is, the help message is general, substituting the local DEFAULT where appropriate. The IF-NEEDED subdivision of the PROBLEM slot, resulting in the automatic display of defaults, is inherited from an even higher-up frame, namely the top-level Procedures frame in the KB hierarchy.

The inheritance mechanism would serve to facilitate the development of a consistent rule base. That is, one might easily "try out" candidate indexing terms (new to the system) by placing them in the existing inheritance hierarchy and thereby finding out which rules would be in effect for them. The MedIndEx KB permits multiple inheritance; in this case, a KB frame inherits all slots from all of its inherits-from parents. It may, however, inherit subdivisions of slots selectively. For example, in general the Bone Neoplasms frame inherits from two frames, Bone Diseases and Neoplasms by Site, but selectively inherits its restrictions (permissible fillers) for the BODY-SITE slot exclusively from Bone Diseases.

The presumed advantages of knowledge-based indexing may be summarized as follows:

- Rules as system-initiated executable procedures
- Inheritance for less redundancy, more rules, less maintenance

- Improved consistency in applying rules
- Document-specific information used for situation-specific assistance
- Domain knowledge available for use in rules

Additional hoped-for advantages of MedIndEx follow:

- Self-training/teaching
Reinforces learning
Conveys rule changes quickly and directly
- Foundation for managing and expanding the indexing scheme
Addresses gaps, inconsistencies, vagueness
- Electronic mail within the system
Facilitates communicating problems to appropriate staff
Serves as electronic suggestion box
- Machine learning from database
Facilitates updating KB classification scheme
Offers users intermediate selections based on probability
- Integration of indexing and search applications using same KB
- Further integration of electronic knowledge sources into the system
- Portability across computer platforms, therefore flexibility as distributed system

The emphasis in this paper has been the "intelligent" features of MedIndEx. However, the prototype also is designed to facilitate data entry and provide many types of user-requested displays. For example, all terms in hierarchical displays, including informative synonyms following the preferred term (see Figure 3), are stringsearchable. If the SCOPE icon is set to ON, the MeSH scope note for a term will appear merely by placing the cursor somewhere over the term in any hierarchical display. A hierarchical display of the entire KB may be requested at any time by mousing on the HIER icon. Multiple display windows can be open simultaneously, subject to a Front/Back toggle. Word-based aliasing with auto-completion is used, e.g., *Est Re Th* is an acceptable entry for Estrogen Replacement Therapy. Permuterm displays may be requested, e.g., *kidney** displays all terms with that word or related words, such as *renal* and *neph-*, including terms having aliases with those words.

Conclusion

We are currently preparing for an evaluation of the system. We hope to have an evaluation design completed by the fall of 1993. The evaluation itself will be aimed at determining the feasibility of implementing a system such as MedIndEx in an operational environment.

References

- [1] Humphrey, S. M. (1992). Indexing biomedical documents: from thesaural to knowledge-based retrieval systems. *Artificial Intelligence in Medicine*, 4(5), 343-71.
- [2] Lancaster, F. W. (1991). *Indexing and abstracting in theory and practice*. University of Illinois Graduate School of Library and Information Science.
- [3] Hutchins, W. J. (1970). Linguistics processes in the indexing and retrieval of documents. *Linguistics* 61, 29-64.
- [4] Slameka, V. (1963). Classificatory, alphabetical, and associative schedules as aids in coordinate indexing. *American Documentation* 14 (3), 223-8.
- [5] Humphrey, S. M. (1989). A knowledge-based system for computer-assisted indexing. *IEEE Expert* 4(3), 25-38.

NASA Scientific and Technical Information

MAI: Maintenance, Performance, Development

CENDI Indexing Workshop

NASA Headquarters

September 21, 1994

**Elliott Linder
NASA STI Program
Information International Associates**

NATURAL LANGUAGE MAI

PROCESS

SELECTS PHRASES FROM TITLES & ABSTRACTS

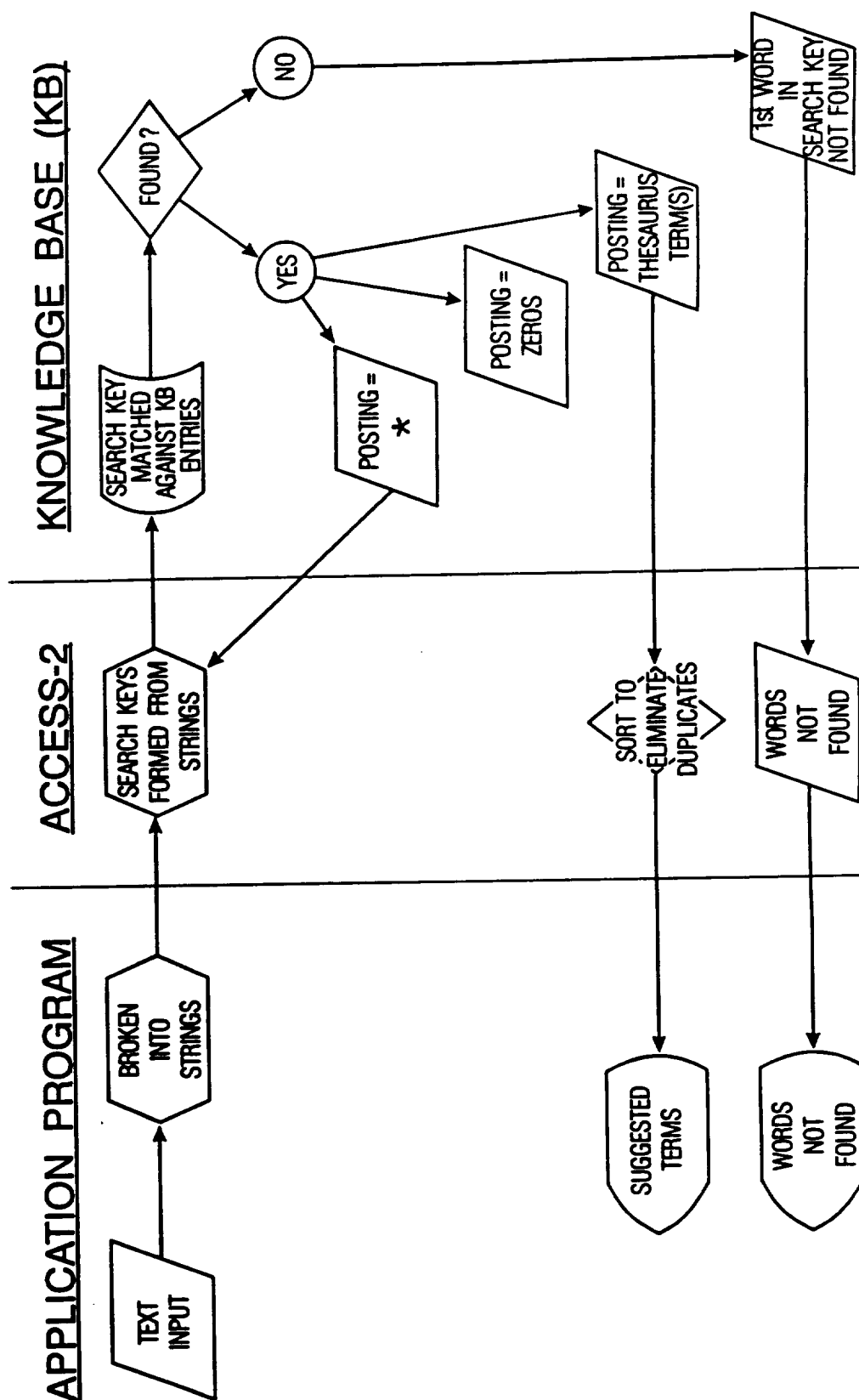
TRANSLATES PHRASES TO NASA THESAURUS TERMS

OPERATIONAL

SINCE JULY 1986

NASA's ONLINE

MACHINE AIDED INDEXING



COMPONENTS OF NASA'S

MAI SYSTEM

- APPLICATION PROGRAM
- ACCESS-2 PROGRAM
- KNOWLEDGE BASE

APPLICATION PROGRAM

- IDENTIFIES INPUT
- SELECTS WORD STRINGS
 - STOPWORDS
 - PUNCTUATION
- PRINTS / DISPLAYS OUTPUT

ACCESS-2

- ACCEPTS WORD STRINGS FROM APPLICATION PROGRAM
- IDENTIFIES SEARCH KEYS FROM STRINGS
- LOOKS UP SEARCH KEYS IN KNOWLEDGE BASE
- SENDS NASA TERMS AND “WORDS NOT FOUND” TO APPLICATION PROGRAM

KNOWLEDGE BASE EXAMPLES

KEY

POSTING TERM

TURBULENT; SHEAR

*

TURBULENT; SHEAR; FLOW

SHEAR FLOW, TURBULENT FLOW

TURBULENT; SHEAR; FLOWS

SHEAR FLOW, TURBULENT FLOW

TURBULENT; WAKE

TURBULENT WAKES

TURBULENT; WAKES

TURBULENT WAKES

TURBULENT; 999

TURBULENCE

TURING; MACHINE

TURING MACHINES

TURING; MACHINES

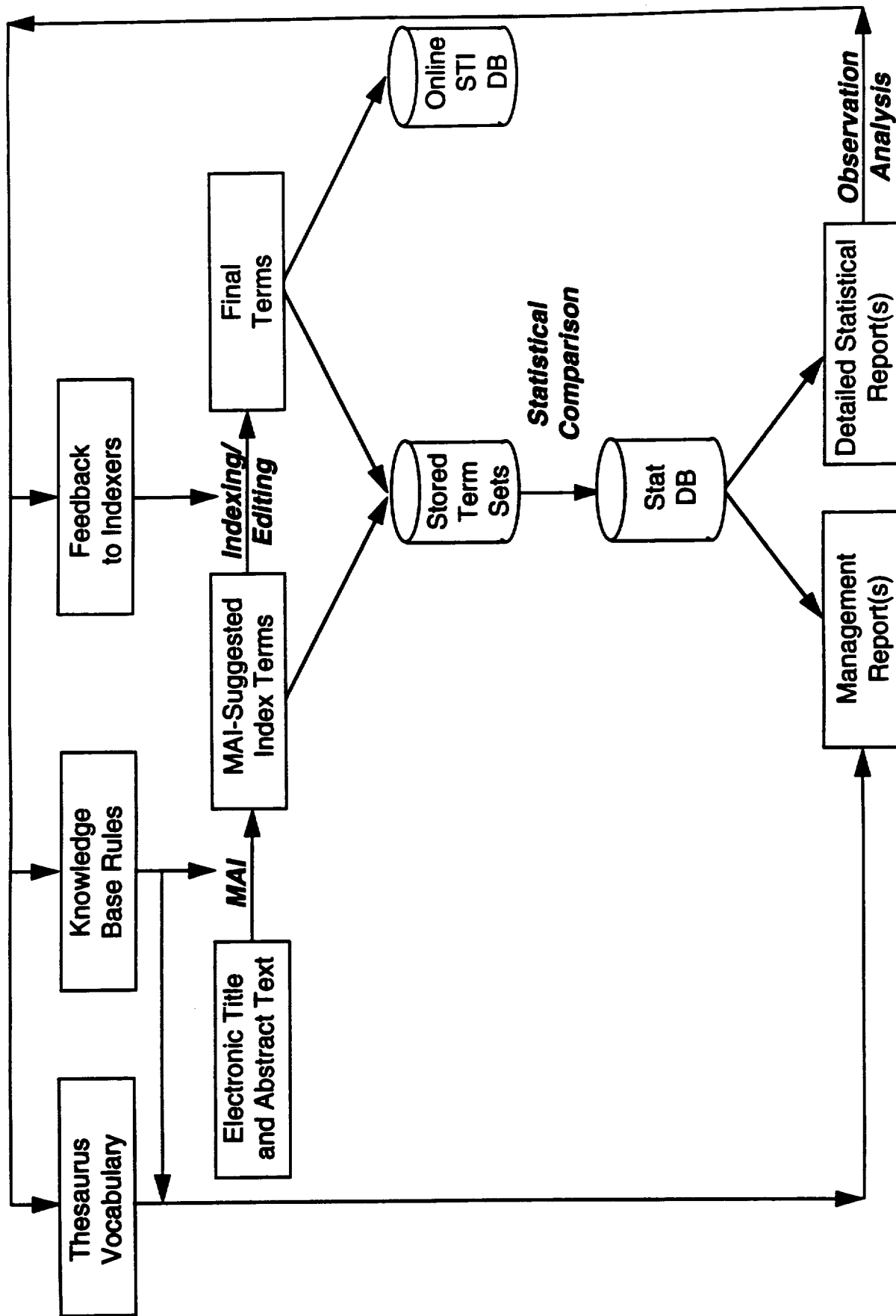
TURING MACHINES

TURING; 999

00

TURION; 999

PLANTS (BOTANY)



MAI Indexing Quality Feedback Loop

**DOE/OSTI Office of Scientific and Technical
Information**

DOE/OSTI INDEXING-RETRIEVAL ACTIVITIES

Charlie Stuber

Office of Scientific and Technical Information

U. S. Department of Energy

615-576-1193

615-576-2865 Fax

charlie.stuber@ccmail.osti.gov

DOE/OSTI INDEXING-RETRIEVAL ACTIVITIES

- **Present Situation**
- **Overview of University of Pittsburgh Studies**
- **Future Directions**
- **Proposal**

PRESENT SITUATION

- **Two-Level Flagging**
- **Decentralized Input**
- **Resource Intensive**
- **Comprehensiveness**
- **Timeliness**

OVERVIEW OF UNIVERSITY OF PITTSBURGH STUDIES

- **Knowledge Base Management**
- **Database Quality**
- **Research In Progress**

KNOWLEDGE BASE MANAGEMENT

- Plasma Physics Subset (18 years)
- Visualization Technique
- Descriptor Lagtime

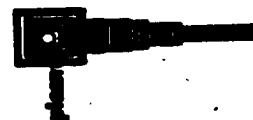


Figure 1: Vibe Point of Interest

horizontal-confinement

.

|

.

|

.

.

fusion

Figure 2: Vibe Document Positioning

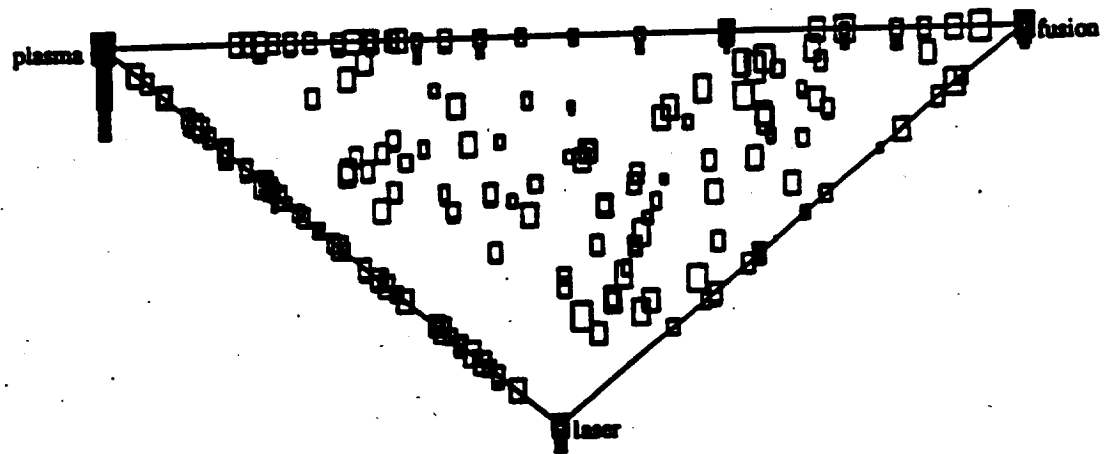


Figure 3. An actual VIBE display with three POIs.

DATABASE QUALITY

- **EDB Statistical Sampling**
- **Citation Validity**
- **Spelling, Grammar, Usage**
- **Indexing Consistency**
- **Categorization**
- **Developed Quality Monitors**

RESEARCH IN PROGRESS

- **Free Text vs Controlled Vocabulary**
- **Use Both; Complementary**

FUTURE DIRECTIONS

- **One-Level Flagging**
- **Free Language With Category Scheme**
- **Thesaurus On Back End**
- **Boolean, Relevance/Similarity Ranking, etc.**

PROPOSAL

- **Category Scheme for Federal (CENDI) R&D and Related Interests**
- **Category Scheme Identifies Commercially Available Databases for Networking**
- **Discontinue Coverage in Agency Databases**

6

Defense Intelligence Agency

Global Change Data and Information System (GCDIS)

The GCDIS is a set of individual agency data and information systems, which provide access to global change data and information.

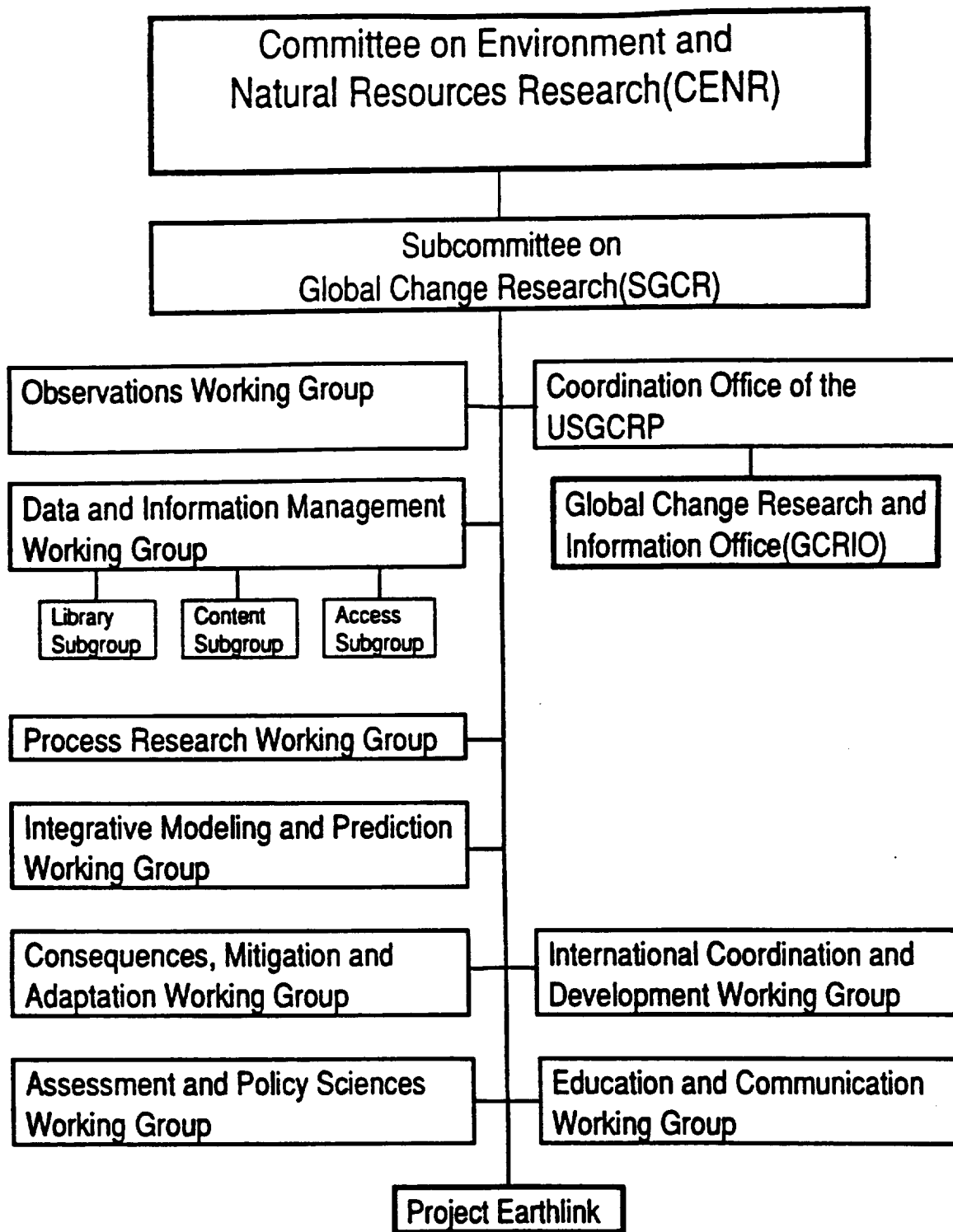
The word "system" embraces the people, hardware, software, networks, telephones, mail, and procedures for identifying, assembling, documenting, archiving and disseminating data and information.

The GCDIS user community extends from global change researchers to other researchers, policymakers, educators, private industry, and citizens. Through the GCDIS the user community will be able to identify and obtain data and information in useful forms for its unique and diverse needs.

GCDIS is responsive to the needs of the U.S. Global Change Research Program. The participating agencies are DOC, DOD, DOE, DOI, EPA, NASA, NSF, and USDA.

Implementation of the GCDIS will be evolutionary, and will depend upon existing agency information infrastructures. Libraries and information centers, both within the federal government and the nation as a whole, provide a gateway for the public's access to global change data and information resources. The issue of public access is especially critical in the area of electronic resources on the Internet which are now available only to a select set of users. Agency initiatives in the development of the National Information Infrastructure are complementary to evolution of the GCDIS, and will undoubtedly enhance and accelerate its successful implementation.

Pilot projects have been and are being developed to demonstrate the uses of the GCDIS and to test the interoperability of the system.



ACCESS TO MULTIPLE DISTRIBUTED DATA RESOURCES within USGCRP Global Change Master Directory

The Need

**More precise and accurate access to research data
and summary information (i.e., metadata)**

The Options

- 1. Keyword elimination**
- 2. Thesaurus development**
- 3. Vocabulary expansion**
- 4. Alternatives to controlled keywords**

A Possible Solution

**"Proof of concept" project to expand the controlled vocabulary
using a semantic network combined with natural language**

The Test

**10,000 documents from participating agencies
ConQuest Software, Inc. commercial available software
6 months**

Findings

**Satisfaction with search capabilities
Limitations in scientific and technical terminology in knowledge base**

Next Step

Assisted Search for Knowledge (ASK) Project

The ConQuest Semantic Network:

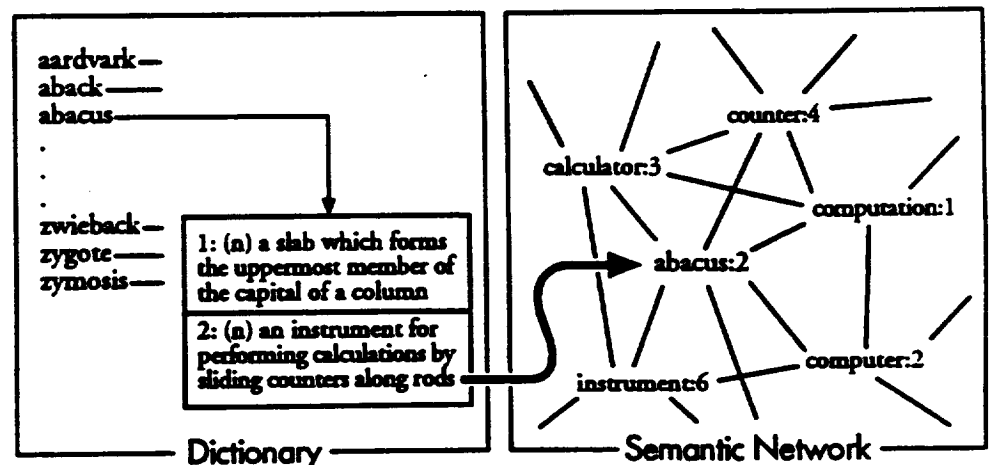
The Power you demand to find the information you need

Plain English ease of use

Enhanced Productivity for full-text applications

ConQuest is the first company to deliver a **Semantic Network** with every copy of its software. Built from published dictionaries and other reference works, this unique resource leverages word meaning and word relationships to find the information you need, with breakthrough accuracy, and plain English ease-of-use. With the Semantic Network, ConQuest automatically retrieves targeted, relevant documents from full-text databases. It transforms information overload into productive knowledge.

ConQuest's baseline network is a knowledgebase of unprecedented breadth and scope: 400,000 word meanings, with 1.6 million links to related words and concepts. Yet, with its multi-layered architecture and dictionary loading tools, it is also the foundation to create flexible, highly tuned knowledgebases for even the most demanding, vertical research applications. We provide a complete migration path from the most general to the most targeted search tasks, for individuals, workgroups and entire enterprises.



The ConQuest Semantic Network

Description

Multiple Sources. Not just a thesaurus, the Semantic Network is built from multiple sources, including full published dictionaries. We continue to enrich the network with specialized reference works for the law, medical, finance, engineering and other disciplines. We can load your proprietary concepts and vocabularies, or provide you with tools to do so yourself.

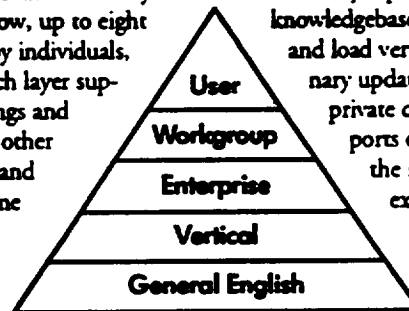
Plain English. Working with ConQuest is as simple as typing a sentence. There is no lengthy set-up time, no complex search language to master. Even ConQuest's powerful Expert mode delivers "point and click" ease-of-use.

Morphology. ConQuest recognizes words at the root level, a much more accurate approach than the simple "stemming" techniques characteristic of other software.

Idioms. ConQuest processes idioms – phrases like "real estate" or "kangaroo court" – not as individual words, but as single units of meaning.

Semantics. No other software accounts for the multiple meanings of words. Simply point and click to the meaning appropriate to your query, for the most accurate retrieval on the market today.

Multi-layered dictionary. A new, multi-layered dictionary structure adds even greater power and flexibility to the ConQuest Semantic Network. Now, up to eight dictionaries can be layered for access by individuals, workgroups, or entire enterprises. Each layer supports over half a million word meanings and can be used separately or linked with other layers. End-users can add definitions and concepts, without affecting the baseline network. Dictionaries can be selected to drive performance for specific tasks, such as technical research or business intelligence.



Dictionary Layers for Enterprise Wide Applications

Benefits

➤ Each source adds strength and richness, in the form of word meanings, word links, parts of speech, word roots and variant forms. This data adds to ConQuest's ability to find an accurate, complete set of information.

➤ ConQuest requires little or no training. Users are up and running in as little as half an hour – spending their time finding and using the information they need.

➤ ConQuest isn't stumped by irregular or variant spellings, and users don't miss out on words which may be relevant to their query.

➤ Recognizing idioms makes for a more accurate search. Rather than false hits on individual words, ConQuest delivers the user right to his/her target.

➤ "Declining stocks" could refer to a bad day on Wall Street or an endangered species. Shouldn't you be able to choose the difference?

➤ The multi-layered structure optimizes the use of dictionary resources, yet protects the integrity of the underlying knowledgebase. Its modularity makes it easy to choose and load vertical dictionaries, and to deliver dictionary updates. Users can even maintain their own, private dictionaries. Its combined capacity supports over 4 million word meanings, breaking the size barrier on what is already the most extensive knowledgebase on the market.

The ConQuest Advantage

ConQuest's automatic 'concept questioning' is transforming the way people access and work with electronic text. It combines out-of-the-box ease-of-use with the power of a full semantic network. It is the only system which allows users to type straightforward, plain English queries, and retrieve documents based on the actual meaning of their query terms.

ConQuest, the leading provider of advanced text management solutions, provides the most complete set of full-text tools available today. Users can index, search, retrieve, browse, categorize and route information, in both archival and real-time applications. ConQuest manages the entire research process, delivering a strategic productivity advantage to your knowledge workers.

Call ConQuest at 1 (800) 787-1715 for further information.



very day, people go to the nation's 115,000 libraries to get information important to their health, employment, financial, and educational needs. Every day more than 13 million people use college and university libraries--more than watch the top-rated TV show. Every year, people borrow more than 1.4 billion items from public libraries.

A free society in today's global and highly technological environment must have access to libraries' collective resources and the expertise of their staffs. For almost forty years, the Council on Library Resources (CLR) has been helping libraries prepare for these vital activities.

Founded in 1956, the Council is a private operating foundation that initiates projects and provides grants. Guided by a diverse board and funded from a variety of independent sources, it has no membership or exclusive constituency shaping its programs. A 501(c)(3) organization, the Council was established to look toward the future on behalf of libraries universally, to address problems experienced by libraries generally, and to identify innovative solutions objectively.

Shaping a Foundation for the Future

Playing a Variety Roles

The Council's independent operating status gives it the flexibility to operate in a manner best suited to the problem it is trying to address. It also has the ability to respond quickly or informally, depending on circumstances. The Council has many roles:

- ★ **Catalyst** Seed money or a letter of endorsement can be all that's necessary to get a good idea off the ground. Many significant library programs today began as small projects initially identified and supported by the Council.
- ★ **Convener** With no vested interest save that of improving library services, the Council can bring together disparate groups of individuals and organizations to address common concerns.
- ★ **Collaborator** Working with other organizations, the Council can bring about significant change. The Council emphasizes cooperative approaches to solving problems and enhancing resources.
- ★ **Facilitator** The Council can focus discussion and encourage projects to expedite operations.
- ★ **Funder** The Council has never had enough money to fund every library's projects, but it does and can support individual projects to resolve problems faced by them all. Council funds have often been used to leverage other institutional resources.
- ★ **Manager** When needed, the Council manages large-scale projects by providing the right mix of administrative and technical expertise.

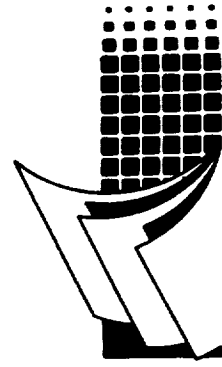
For more information about the Council's programs or funding requirements, please contact:

COUNCIL ON LIBRARY RESOURCES

1400 16th Street, N.W., Suite 510
Washington, D.C. 20036
Tel. (202) 483-7474
Fax (202) 483-6410
Internet clr@cni.org

A Look at...

The Council on Library Resources



The Council Library Resources has a history of funding landmark projects that have left their mark on the national and international library scene. It works closely with major research libraries and library organizations.

Among its earlier achievements is its support for building the infrastructure that enables cooperative library activities. For example, the Council coordinated efforts in the 1980s to permit the linking of online library catalogs. Researchers can now search the collections of libraries across the country from a single location. A brief history of the Council's contributions is found in our booklet, "Shaping A Foundation for the Future." A list of current projects is also available.

Helping Libraries Prepare for Tomorrow

Libraries of today benefit from the many projects supported by the Council. The fact that our nation's libraries are part of an international information system can be attributed partly to the behind-the-scenes programs and projects of CLR.

The Council sees the future library as an electronic window to a world of information. As information increasingly becomes available through new forms of publication, libraries continually strive to improve access to that information. In order for libraries to be able to respond effectively to their users' needs, they must use library resources judiciously.

Focusing on Library Resources

As today's libraries made the transition to automation, they benefited from the

advice and experience derived in the Council's trailblazing projects. Tomorrow's libraries will face new challenges. To prepare libraries for their tomorrow, the Council is committed to the development, sharing, and innovative use of library resources. Examples of those resources are described below.



People

More than 182,000 professionals work in libraries today. As new technologies increase the variety of means of access to information, the demand for the skills of librarians and other information professionals will continue to grow. New skills will be needed as well. The Council's human resources program examines how we attract, educate, and advance individuals in the information service profession. A major focus of this program is to assist current leaders to develop the skills needed to transform their institutions in response to societal needs. Grant support from the Council helps librarians develop their research and leadership skills; CLR also has supported intern and mentor programs.



Collections

College and university libraries in the United States contain over 717 million volumes. Public librarians answered 222 million reference questions in 1991. The amount and type of material made available in published and unpublished form each year overwhelm the selection and reference functions in our nation's libraries. Through its economics program, the Council seeks mechanisms to assure that the financial resources invested in libraries and related information services will best meet the needs of users.



Space/Buildings

In any given year, there are over 700 library building projects in progress. Each year more than 250 such projects are completed, accounting for over

\$860 million dollars in expenditure. The Council's infrastructure program explores how new technologies will transform the structures that house our libraries as well as the information systems and services that support library activities. Infrastructure issues extend beyond buildings and systems to assure that economic, sociopolitical, technical, and legal changes do not inhibit library function or access to information by individuals and groups.



Systems/Networks

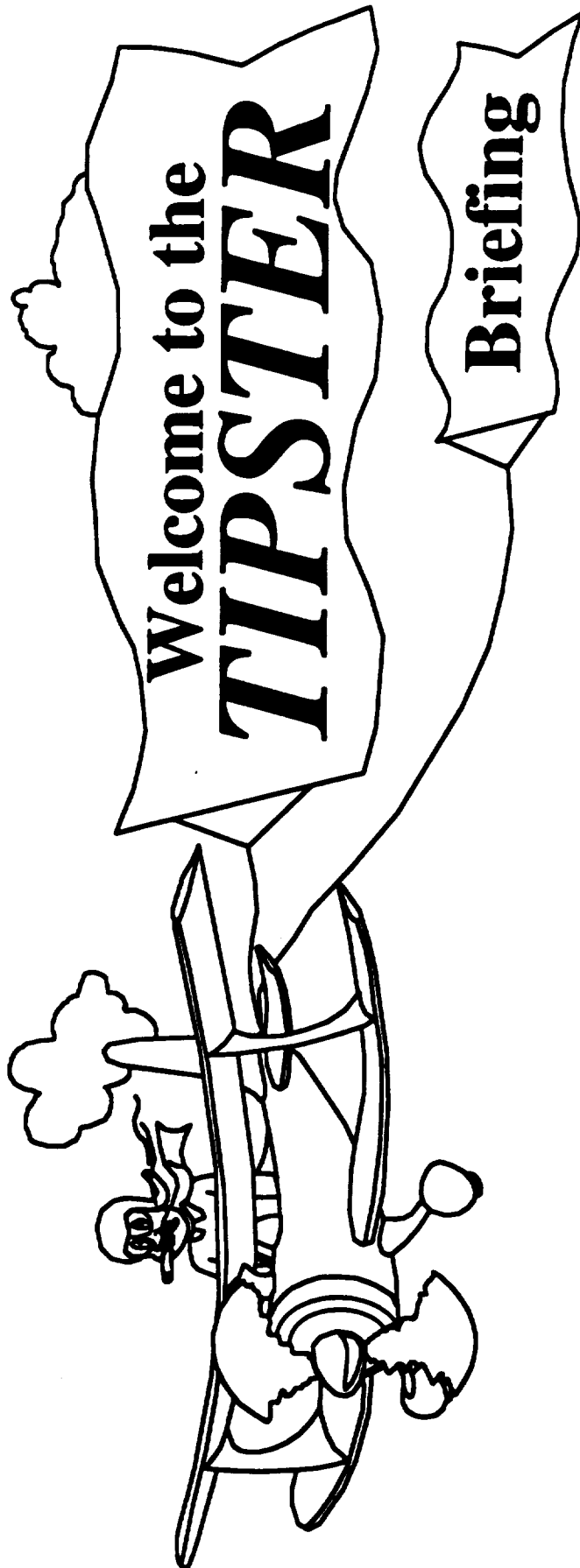
Seventy-five million Americans are computer users; more than 13 million of them have access to online information. Through its access and processing program, the Council has supported the development of standards for libraries to share information across systems, and continues to support activities that address policy and technical issues. Libraries must continue, as they have in the past, to look for appropriate new technologies to improve the methods by which information is acquired, organized, stored, retrieved, reproduced, and made available for efficient use by the communities they serve.

Benefiting Generations of Citizens

The Council on Library Resources was established to address the problems of libraries generally. Our history demonstrates our interest in finding creative solutions to help libraries improve their operations. With each decade some problems are solved and new ones emerge. Since times and technologies change, our work is never done. We are needed now more than ever. Current and future generations of citizens will benefit most from improvements in library services. Libraries and their resources are strategic to the fabric of any society, and they need a strong foundation upon which to build. The Council on Library Resources will continue to serve them well.

7

RMS Associates



#1

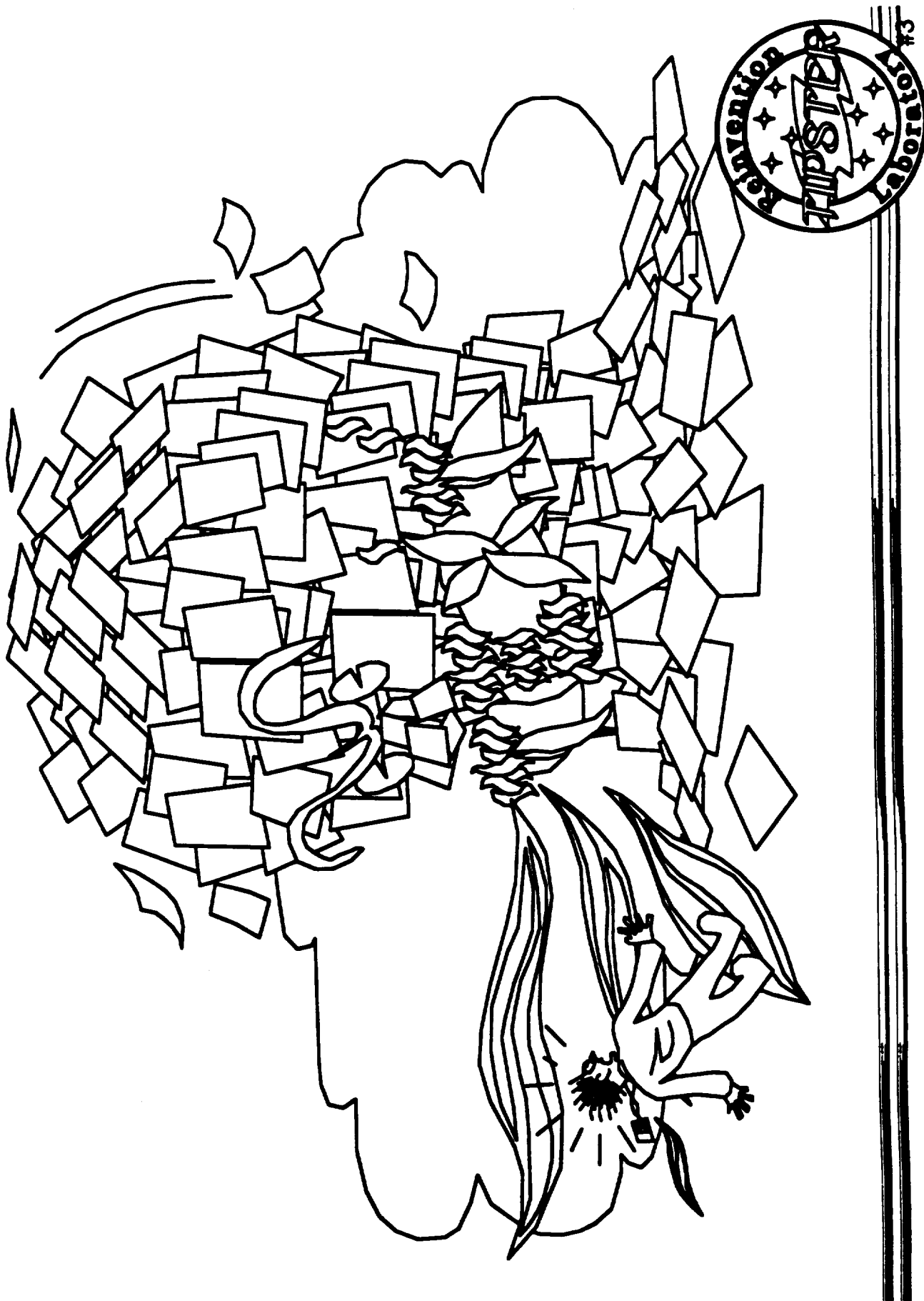
and the wall
came tumbling
down.

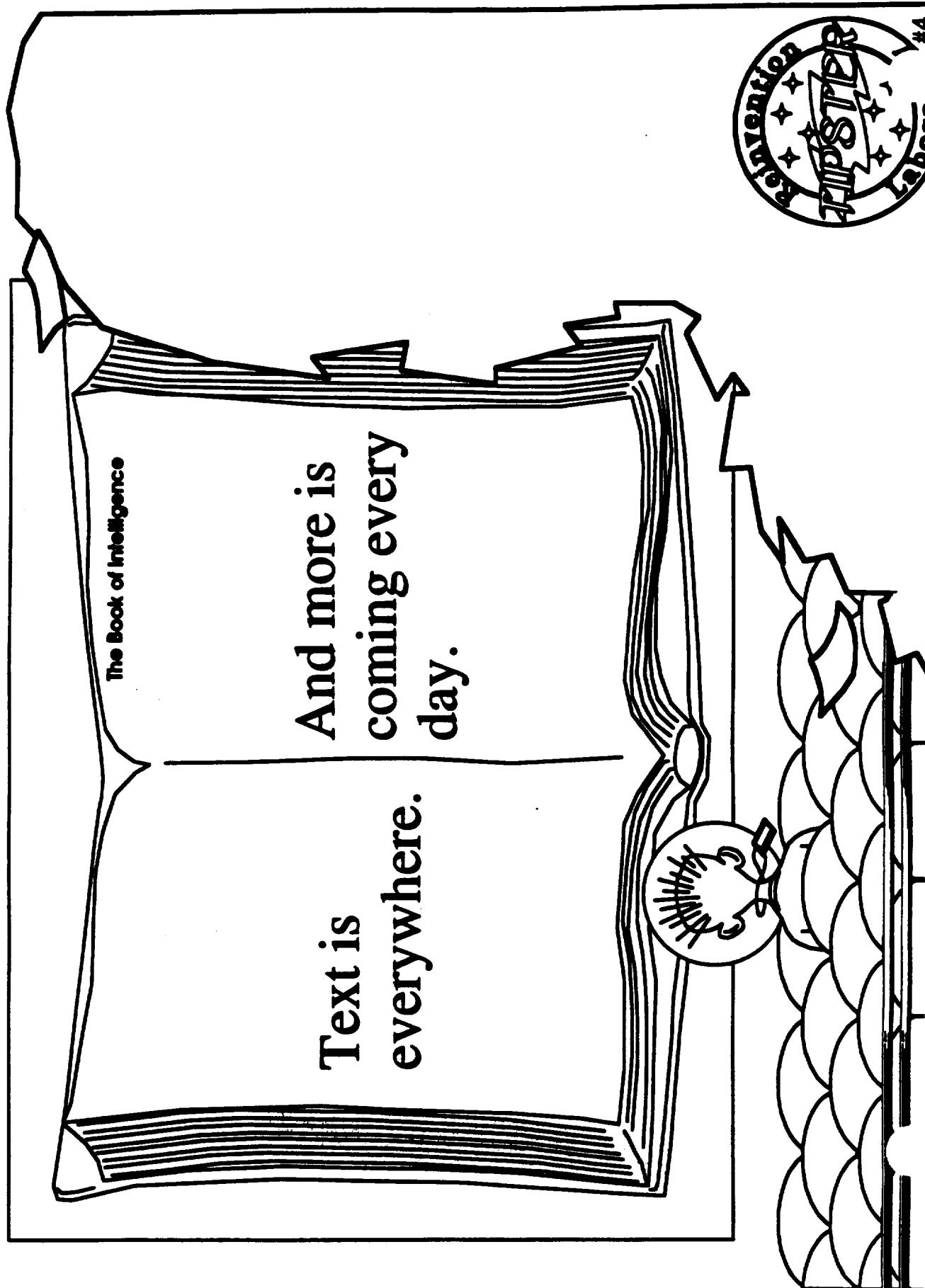
The Book of Intelligence

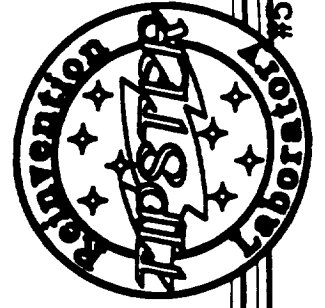
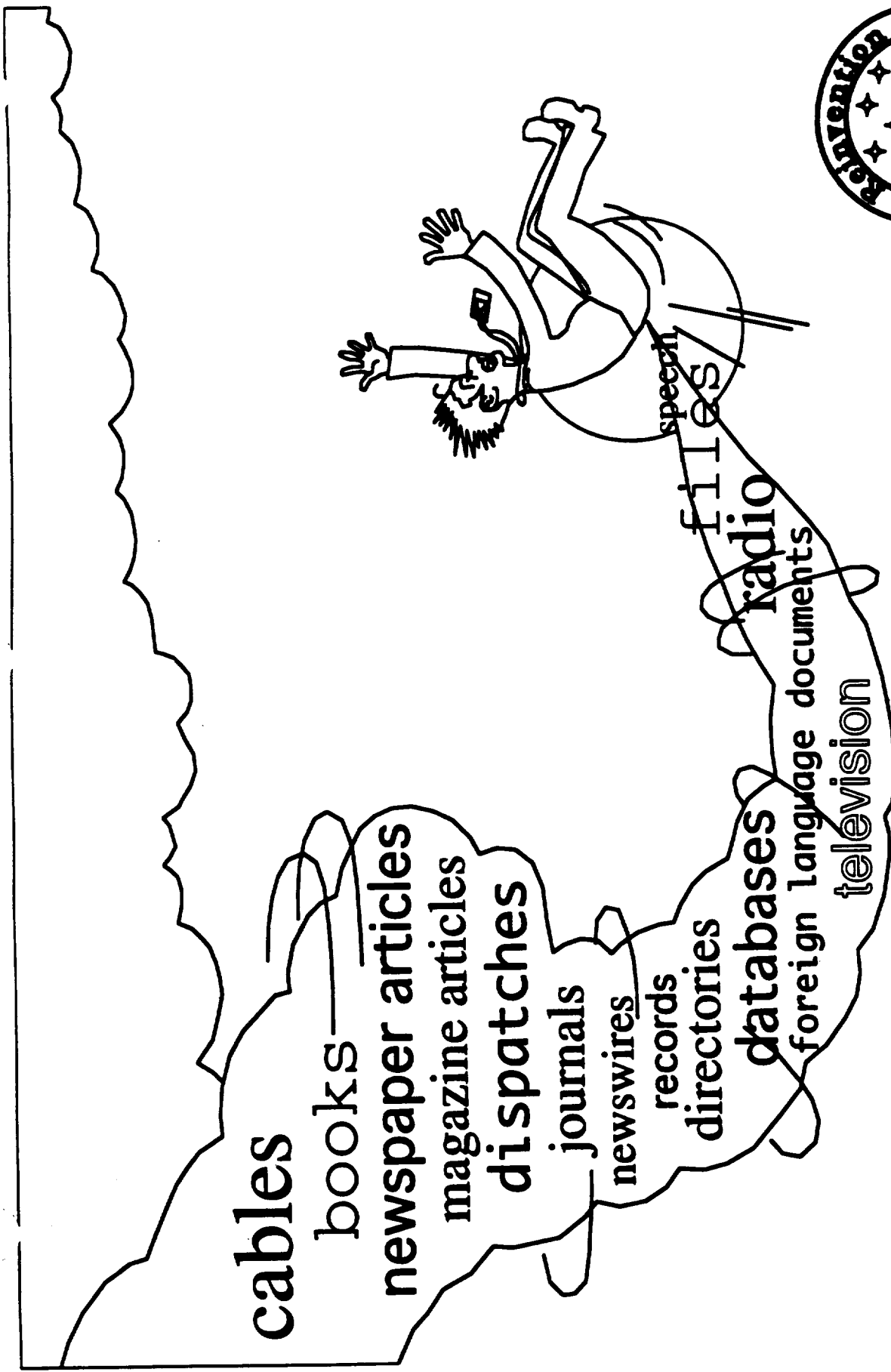
Chapter X

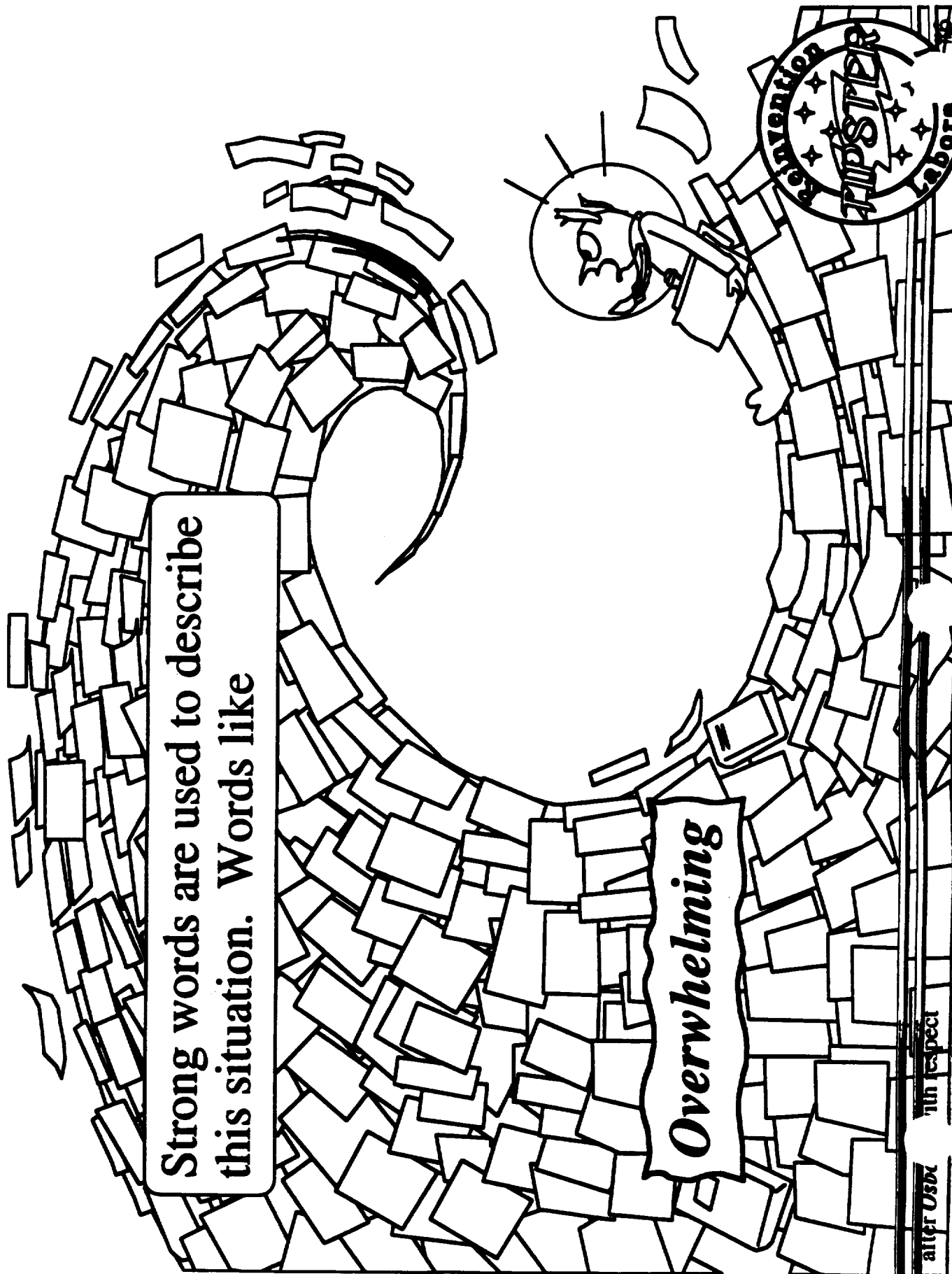
Perhaps the
biggest
problem in the
intelligence
world today is

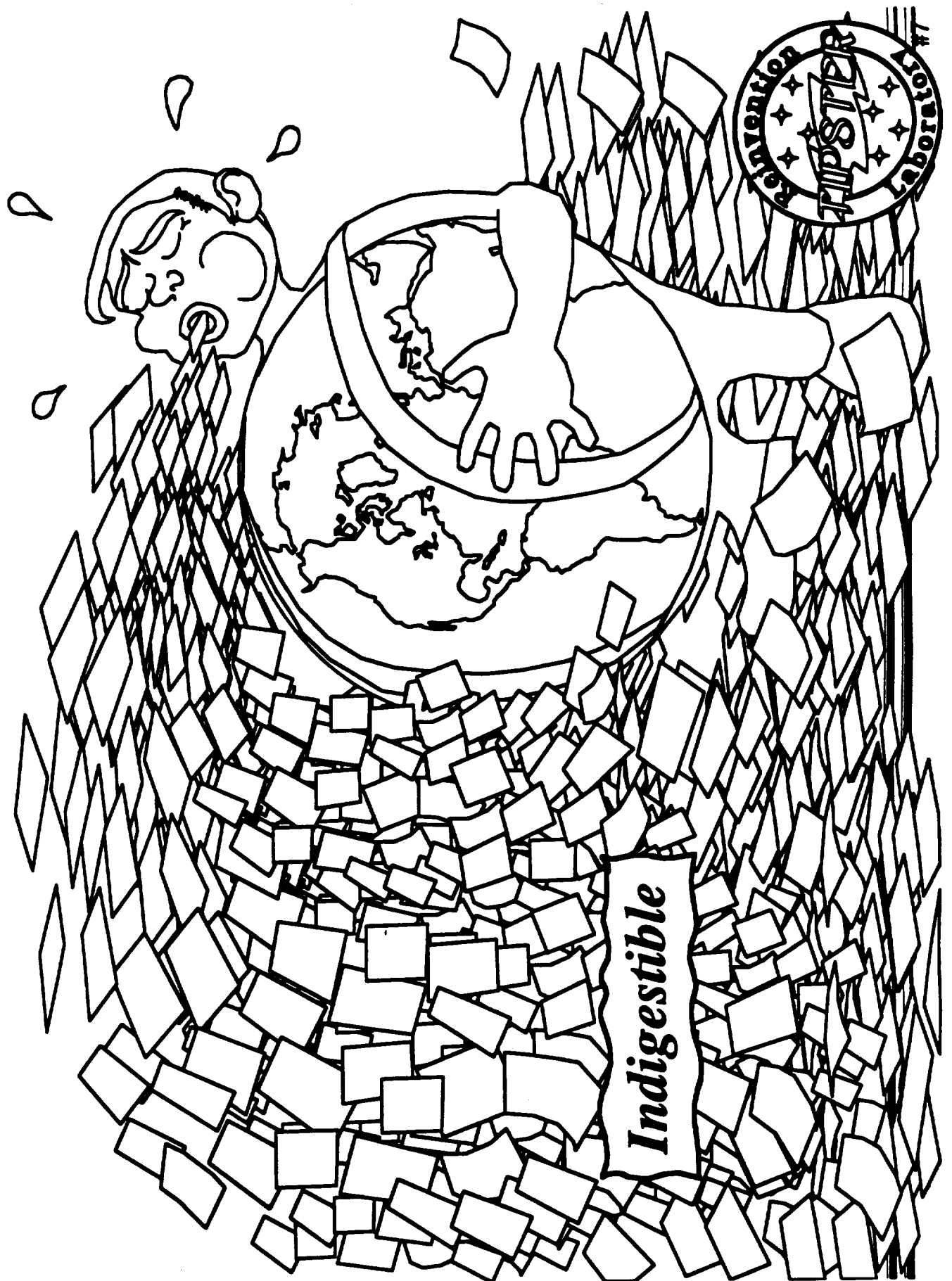








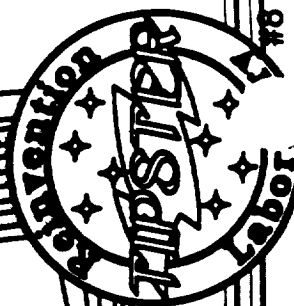




The Book of Intelligence

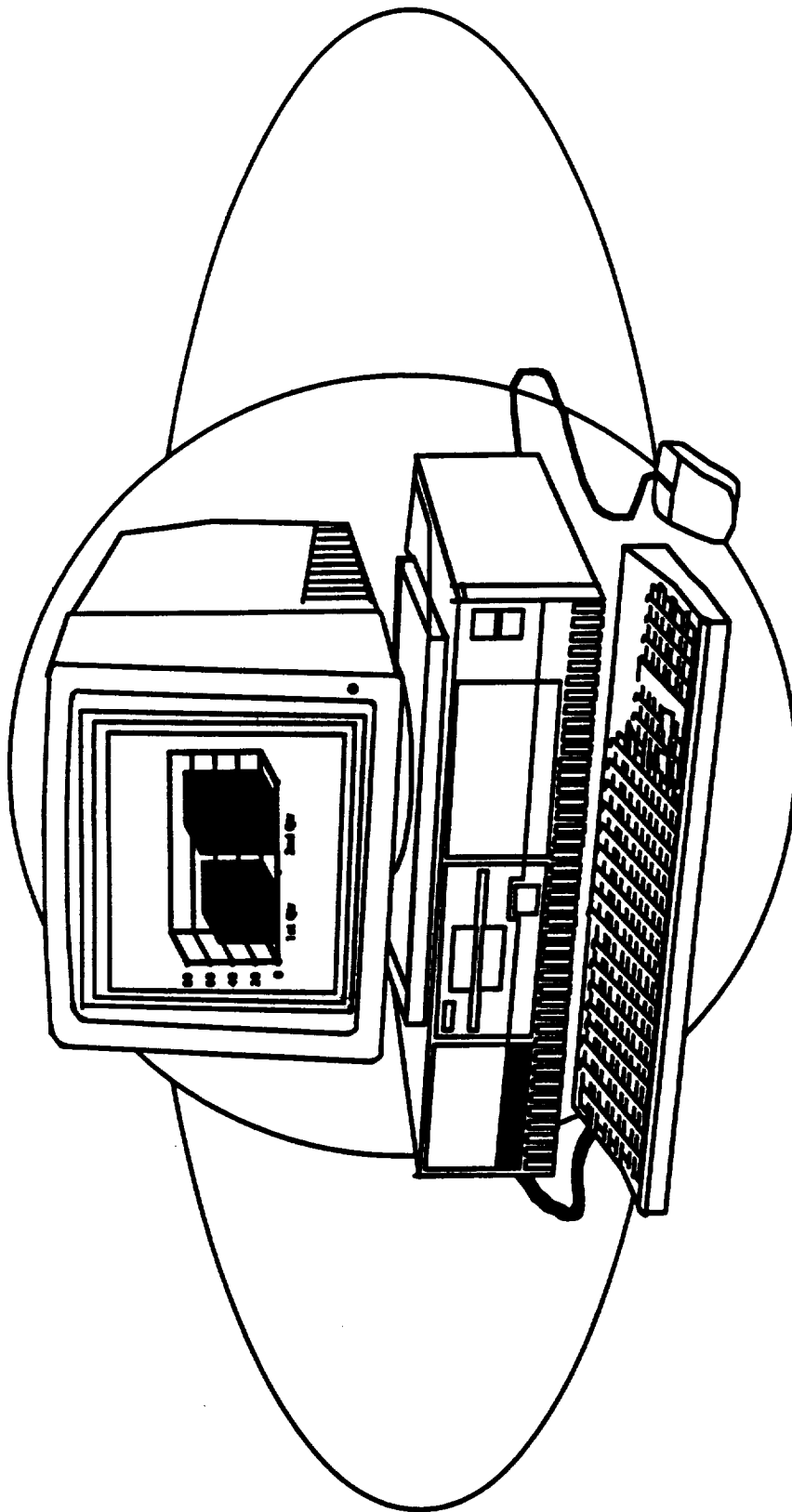
Capturing text
in machine-
readable form
is a difficult
procedure.
But it pales in

comparison
with the major
problem: the
analyst's
*requirement for
information...*

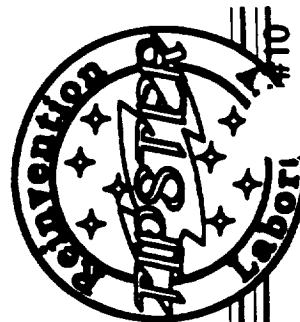




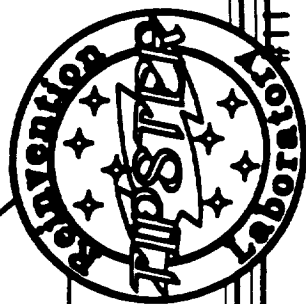
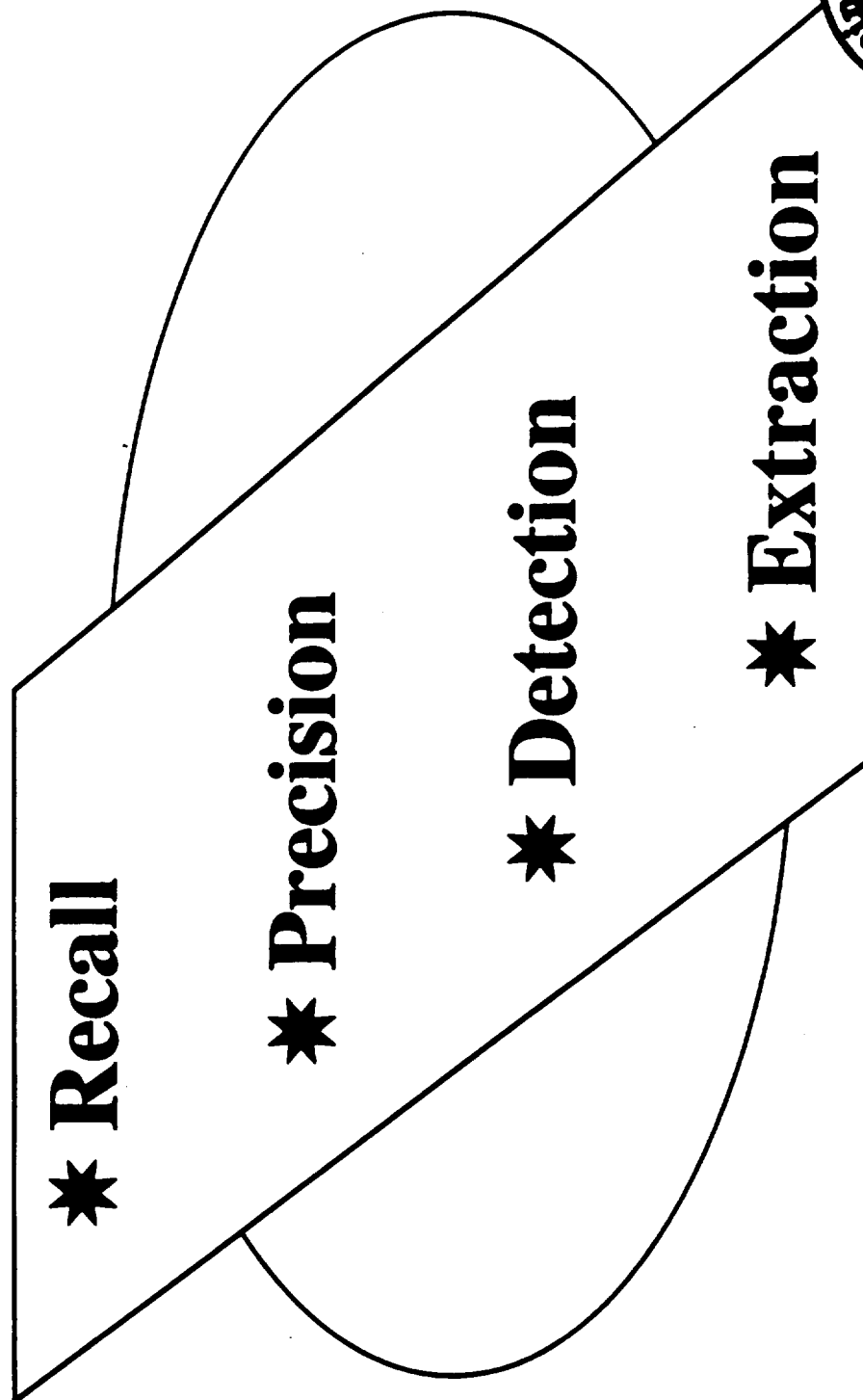
The computer is the only hope to solve this problem

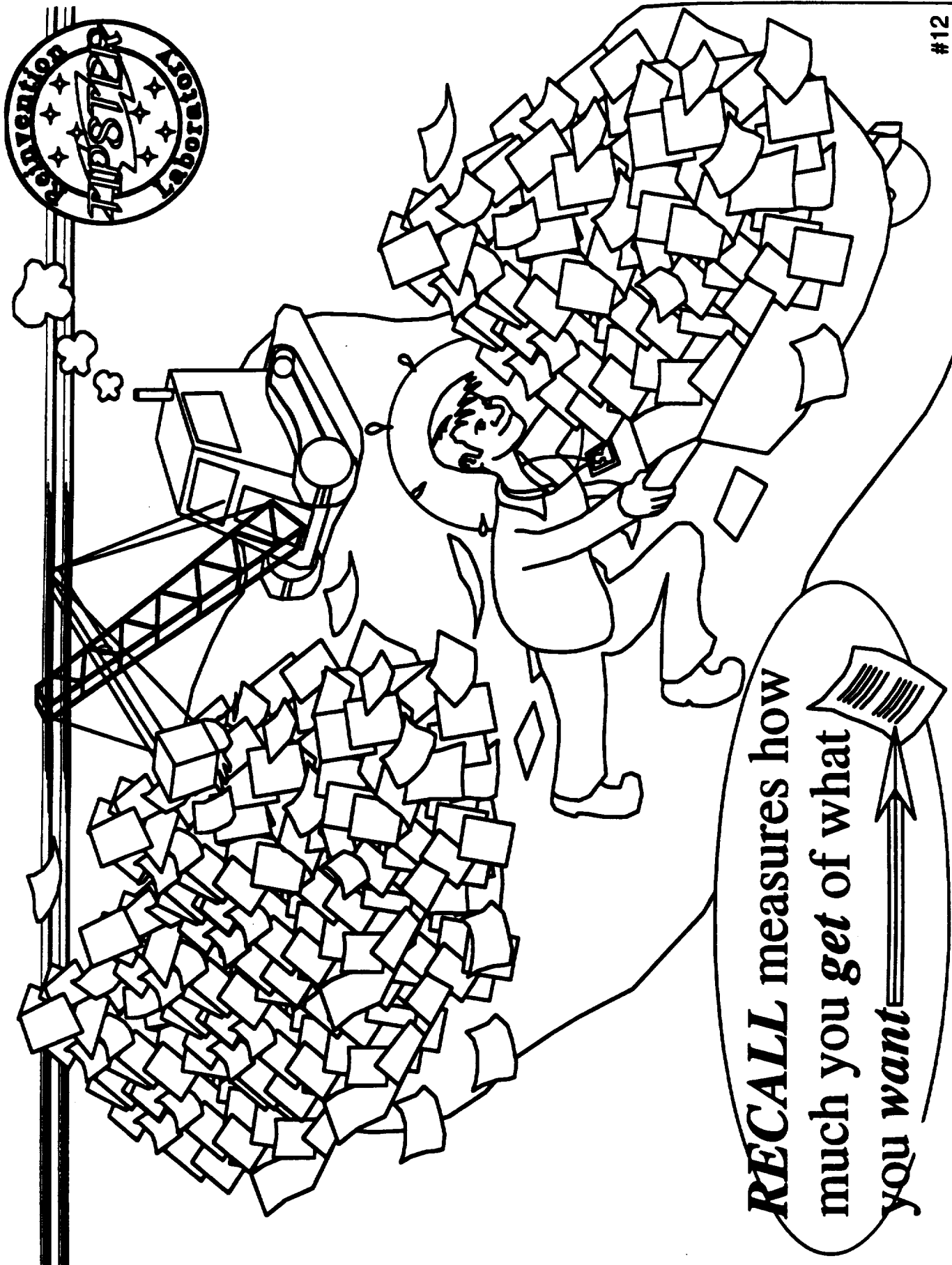


but progress has been slow -- and no commercial software satisfies the need



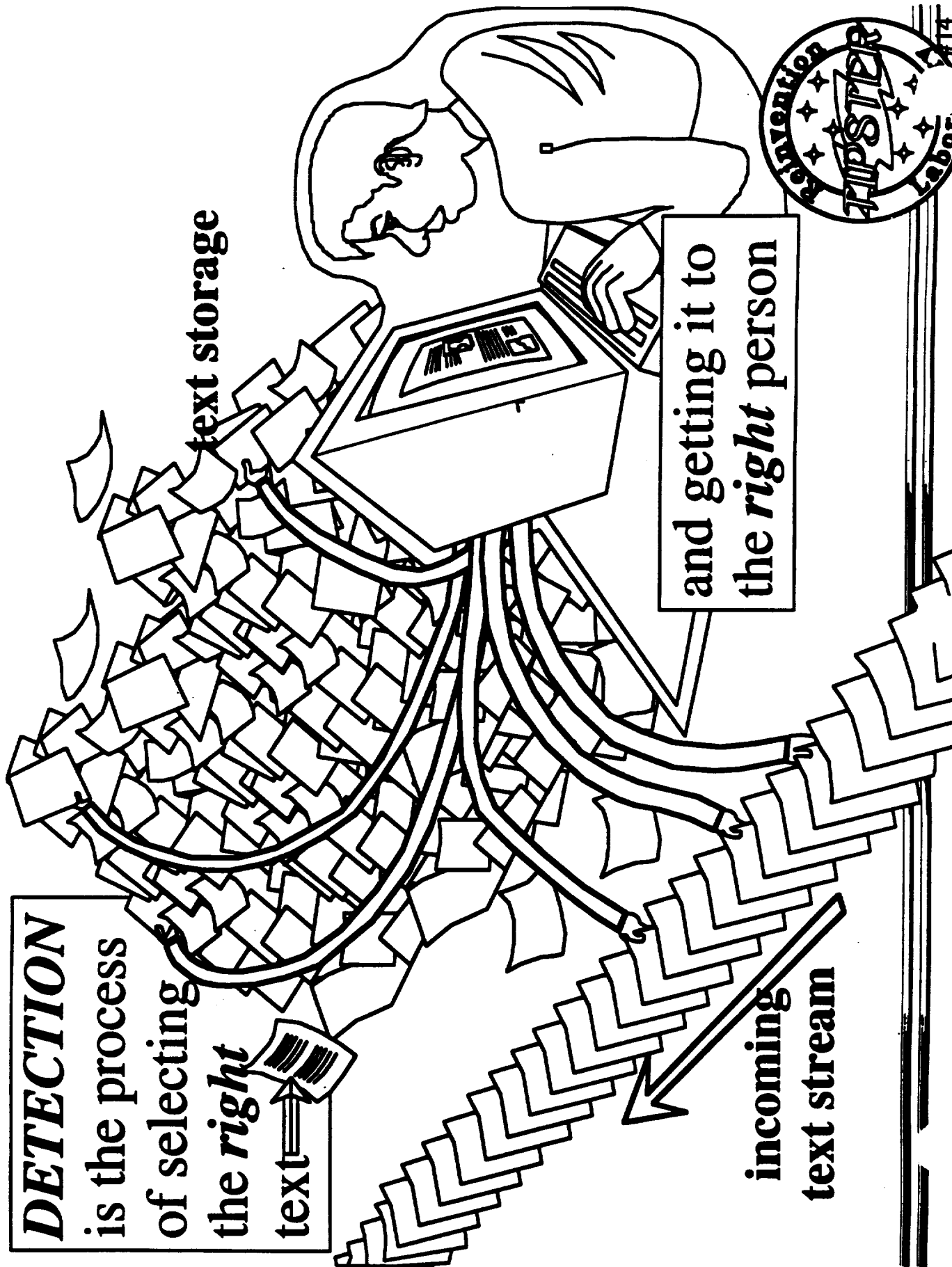
The field of *information selection* uses these terms



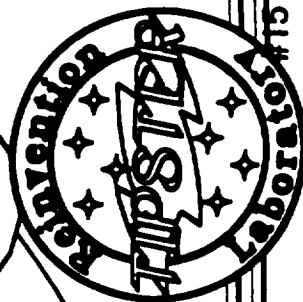
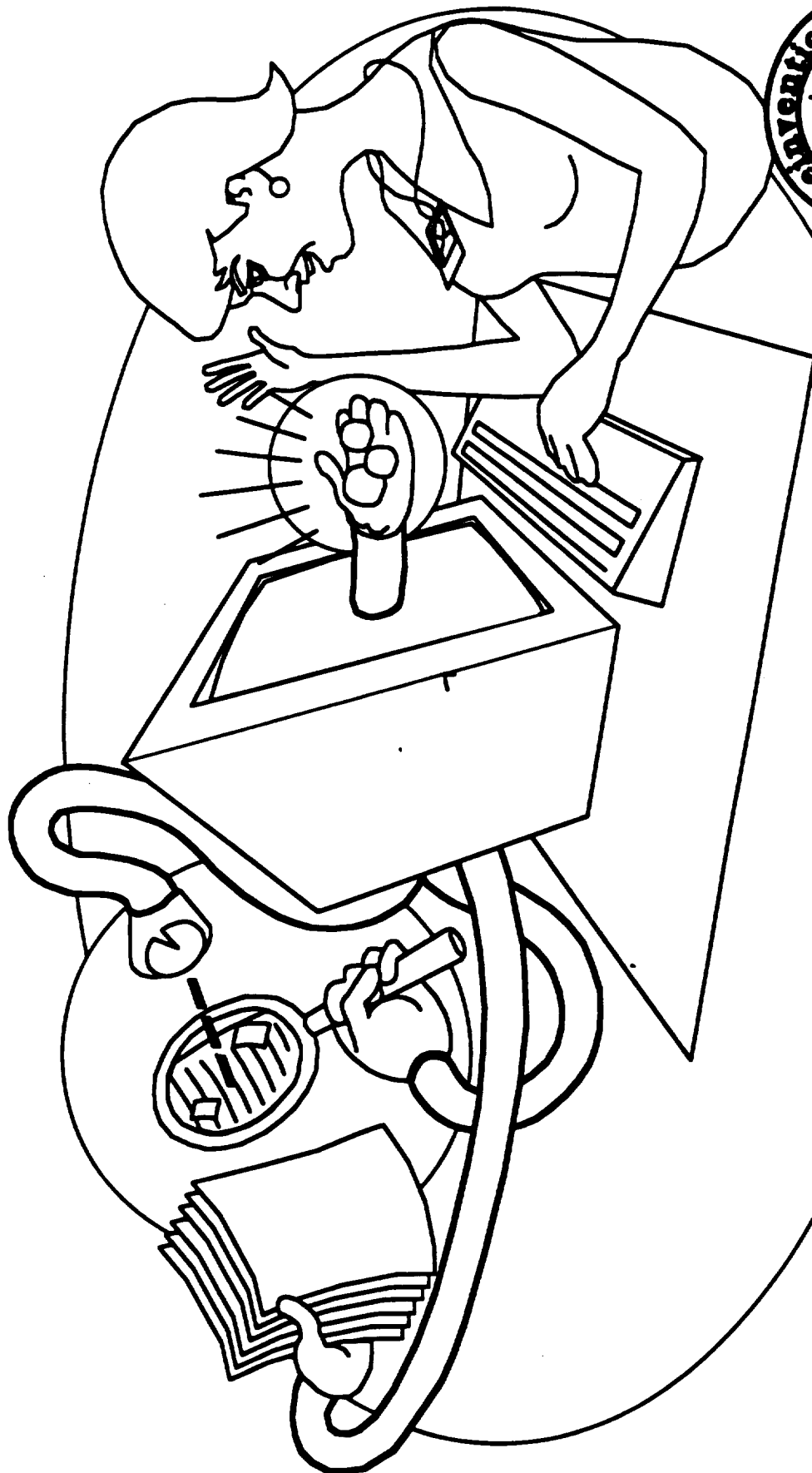


PRECISION measures how much
you want  of what you get





EXTRACTION distills the text down to its essence



Urgency of the Challenge

To exploit text today analysts must

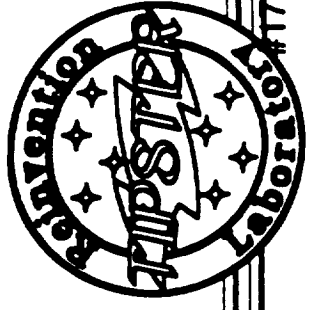
- Work with 25-year-old technology
- Be grateful for 20% recall rate
- Use English-*only* tools
- Use backlogged databases, or none at all
- Do without most available analytical tools



Urgency of the Challenge

Three years ago, no relief was in sight

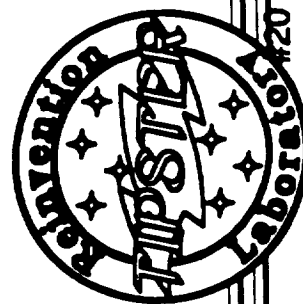
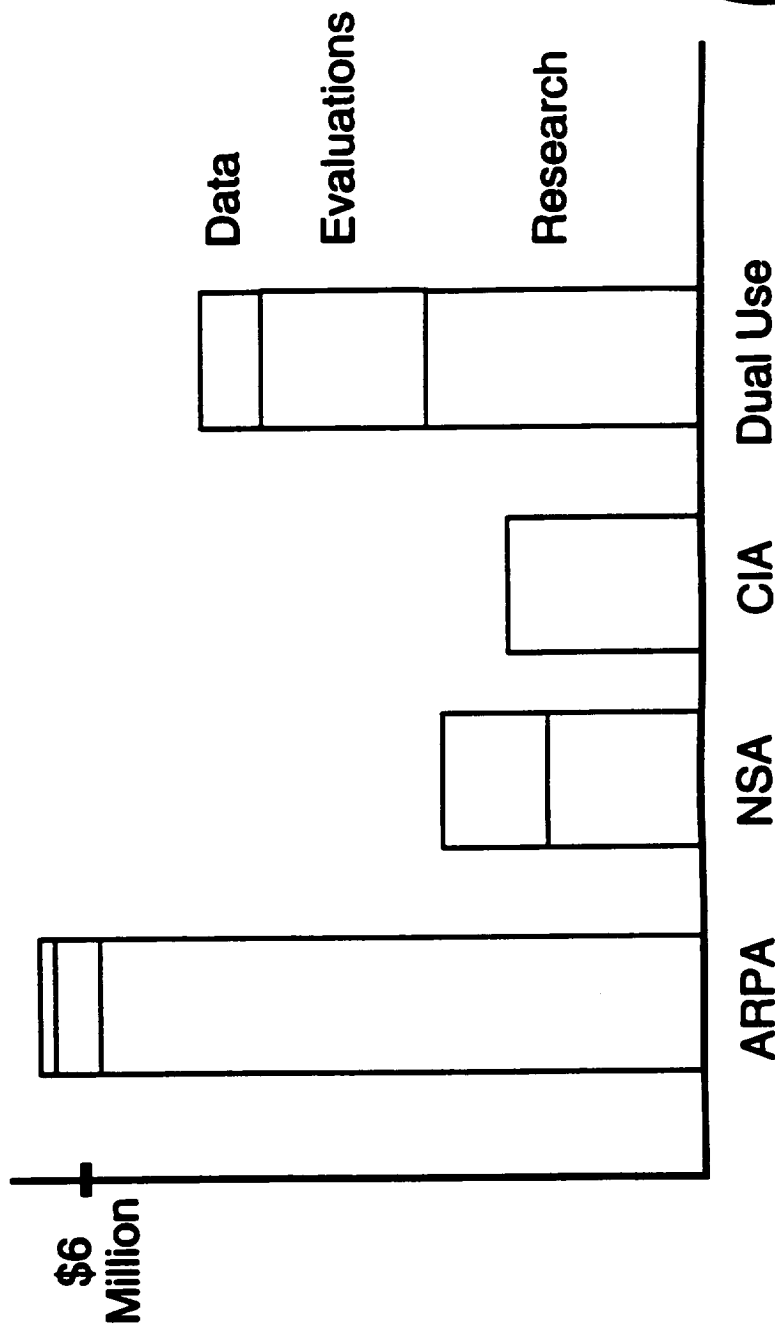
- **Detection: performance was inadequate**
- **Data extraction: quality was "researchy" and brittle**
- **Recall and precision: little market pressure to increase**
- **Research community: stagnated and focused on ivory-tower issues**





Funding: Sources and Expenditures

Tipster Phase I total: \$15.6 Million



Program Timeline

Phase I

 **Planning**

BAA Published  **Contract selection**

 **Data preparation**

4 document detection algorithm
research contracts

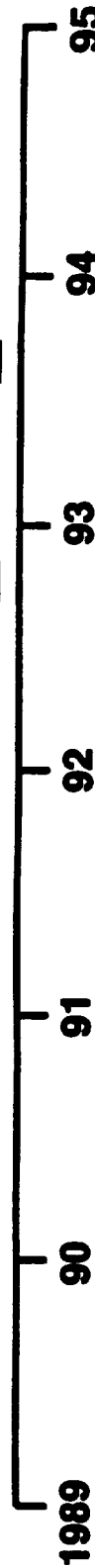
4 information extraction algorithm
research contracts

TIPSTER program workshops 

DEA information extraction system (Project HOOKAH) 

MUC-3  **MUC-4**  **MUC-5** 

TREC-1  **TREC-2** 



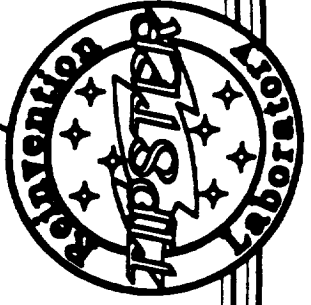
Phase I: Contractor Participation

- **DETECTION**

- University of Massachusetts/MCC - Inquiry (NSA/P043/R52)
- HNC, Inc. (Hecht-Nielson) - Match Plus (CIA/ORD)
- Syracuse University - Dr. Link (CIA/ORD)
- TRW - JFDF/Sleeper (NSA/P043)

- **EXTRACTION**

- General Electric/Carnegie Mellon University - Shogun (CIA/ORD)
- Bolt, Beranek and Newman, Inc. (BBN) - Plum (NSA/P043/R52)
- New Mexico State University/Brandeis University - Diderot (NSA/R52)
- University of Massachusetts/Hughes Research - Circus (NSA/R52)



Commercial and Academic Participants

CORE TIPSTER

DETECTION

HNC

Syracuse

TRW

University of Massachusetts

EXTRACTION

BBN

Brandeis

Carnegie-Mellon University

GE-Corporate Research

Hughes

Martin-Marletta

New Mexico State University

University of Massachusetts

SUPPORT

DynCorp Meriden

IDA

SAAC

SAAC-SETA

MUC-5

AT&T

BBN

Brandeis

Carnegie-Mellon University

Columbia University

DSR

EE-Systems

ETL

GE Corporate Research

Hughes

IDA

LSI

Martin-Marletta

Mitre

NEC

New Mexico State University

NYU

Pacific Sierra Research

Paramax

PRC

SAAC

SRA

SRI

Sterling Software

TRW

Unisys

University of Manitoba

University of Maryland

University of Michigan

University of Sussex

TREC-2

ADS/Booz Allen & Hamilton

AT&T

BBN

Bellcore

Carnegie-Mellon University

Clarit

City University, London

Coherent Research, Inc

ConQuest Software

Cornell University

Dublin City University

ERIM

GE Corporate Research

HNC

IDS

Information Systems

Martin-Marletta

Meed Data Central

NEC

NYU

Paracel

PHC

Queens College, CUNY

Royal Melbourne Institute

Rutgers

Siemens Corporation Research

SLIS, University of California

Syracuse University

Systems Environment

Corporation

Thinking Machines Corp

TRW

University of California

University of Cambridge

University of Central Florida

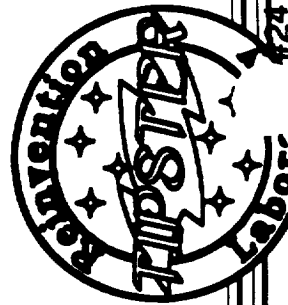
University of Dortmund

University of Illinois

University of Massachusetts

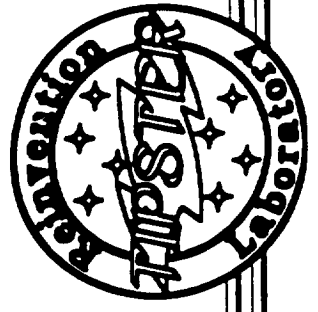
Verity

Virginia Polytechnic Institute



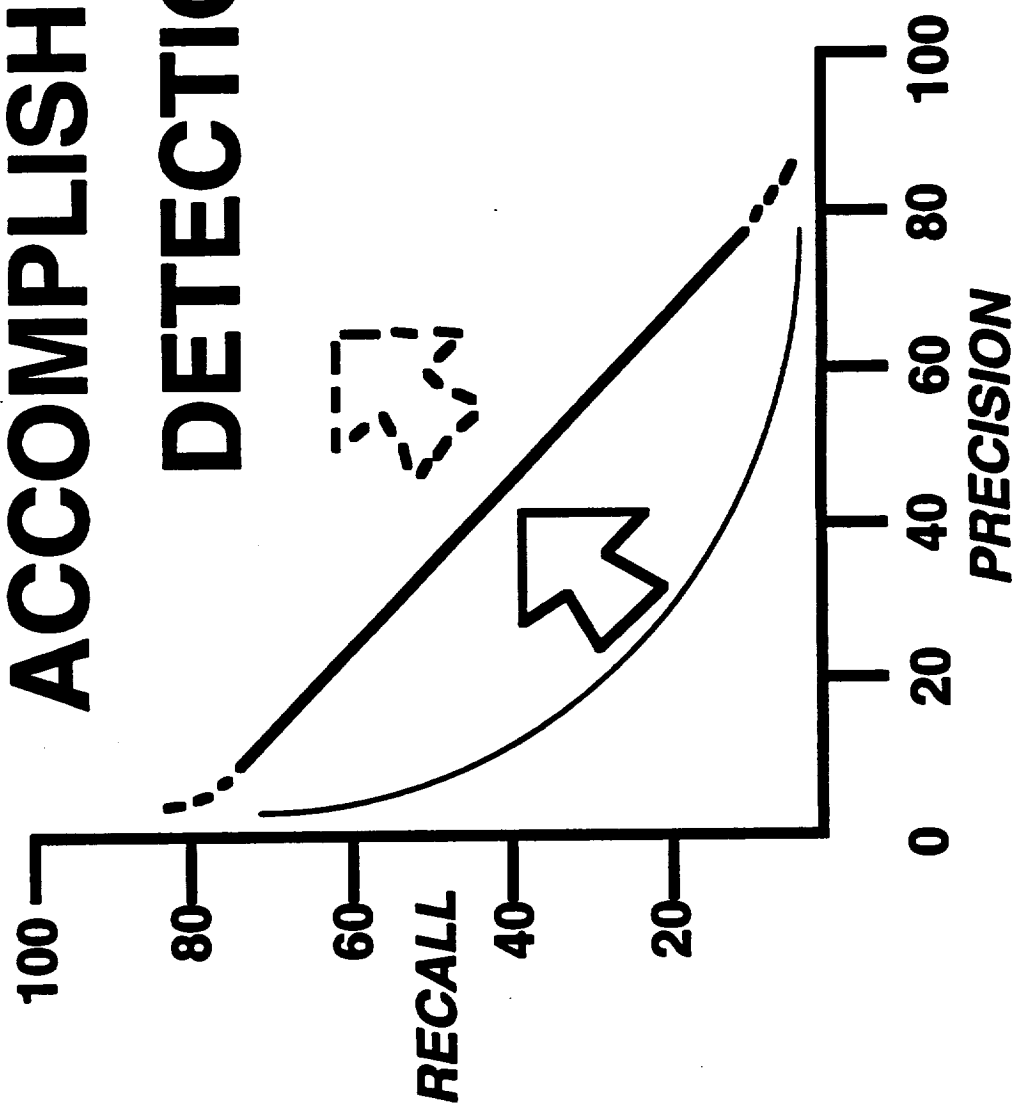
Phase I Accomplishments

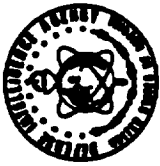
- Technology ready for operational testing
- Major new corpus reflecting government research and development priorities
- Proven mechanism for influencing the research and development communities
- Immediate impact on government programs and on the commercial sector



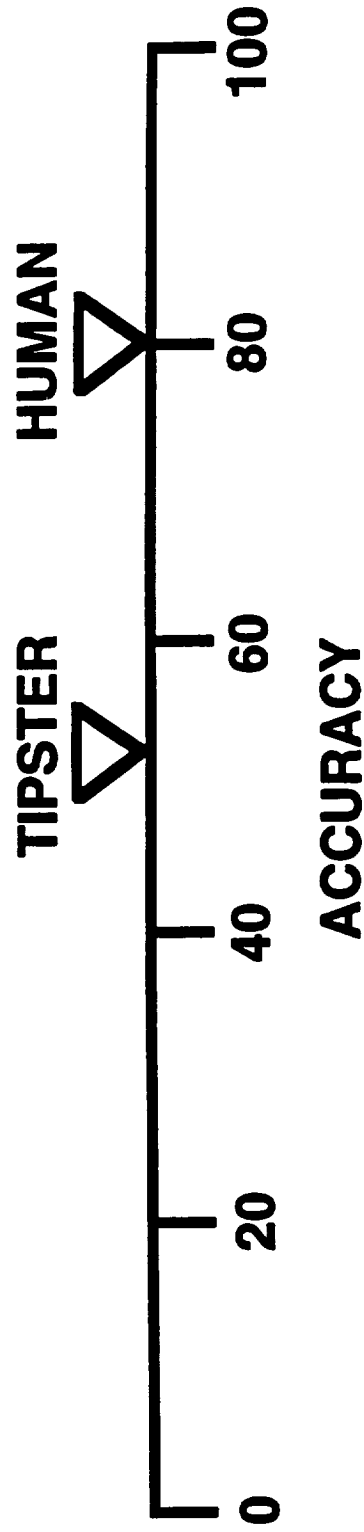


PHASE 1 ACCOMPLISHMENTS DETECTION





PHASE 1 ACCOMPLISHMENTS EXTRACTION

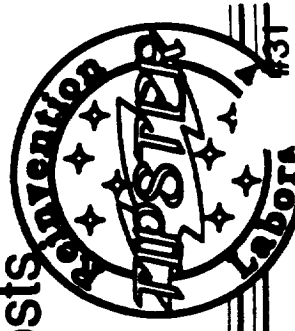


01M9369-2/BKK

Phase II Summary

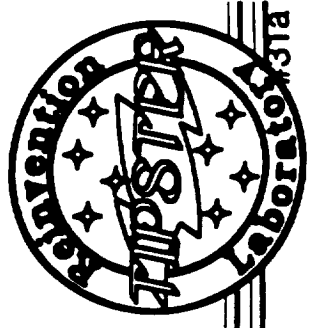
Goals of second phase:

- Maintain research leverage by continued interagency cooperation
- Foster further research on IC issues by continued support of evaluation conferences
- Continue selective support of advanced research
- Fieldtest promising technologies
- Design for reduced acquisition and life-cycle costs



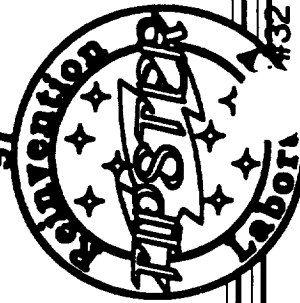
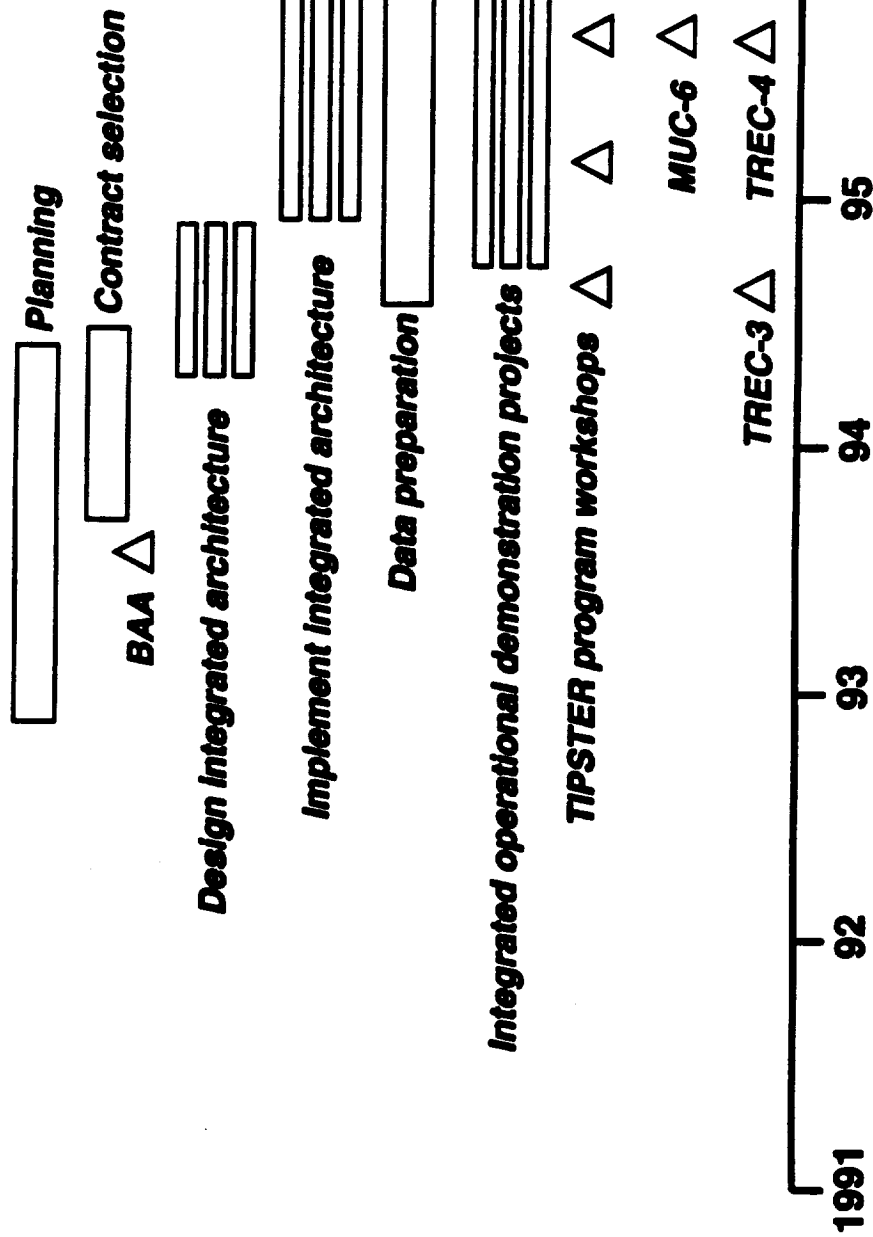
Phase II: Contractor Participation

- **Architecture and Research**
 - BBN, GE/CRD, HNC, Martin Marietta, NMSU, NYU, SRI, TRW, UMass
- **SE/CM and Program Support**
 - 'PRC
- **Demonstration and Example Projects**
 - BBN, Martin Marietta, and TBD



Program Timeline

Phase II



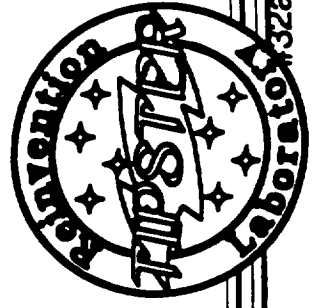
Phase II: Architecture Effort

- **Purpose**

Provide a vehicle to deliver Tipster technology to the workplace flexibly and with manageable cost

- **Goals**

- Combined detection and extraction
- Multiple human languages
- Appropriate response times
- Modular substitution
- Maximum platform transportability
- Scalability to wide variety of text volumes
- Compatible with multi-level security



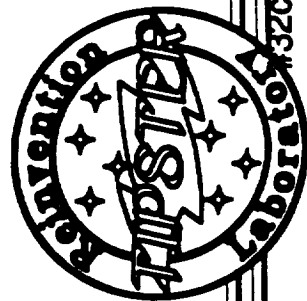
NSA Demonstration Projects

<u>office</u>	<u>description</u>
B Group	Foreign language detection and extraction of person/organization information
B Group	Foreign language detection and extraction of information for input to a tracking system
B Group	Detection and extraction of economic information
W Group	Multilingual document detection and extraction of science and technology information
DDO	Retrospective retrieval from massive, highly heterogeneous on-line text database
SOLIS	Front end processing and reformatting; retrospective retrieval of analyst published reports



CIA Demonstration Projects

<u>office</u>	<u>system</u>	<u>description</u>
FBIS	PRIDES	Retrieval and routing of FBIS product using HNC or UMass product
OSPO	ADEPT	Capture and format of open source text using BBN or GE/Martin Marietta product
-----sensitive-----		Automated support for message indexers using GE/Martin Marietta product

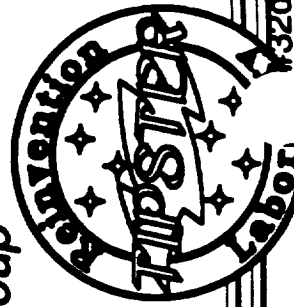


DIA Demonstration Project: PATRIOT

Pilot Application of Tipster for Corporate Indexing of Text

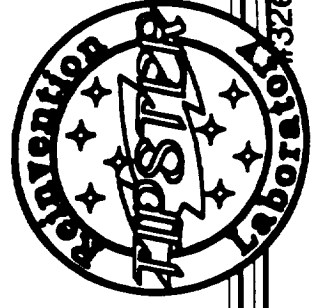
Automated support for cataloging (detection, prioritization, bibliographic indexing, topic indexing) and “deep indexing” (database fill of entity, relationship, and event data) of all-source text using one or more detection and extraction products.

Note: A corporate indexing group and an analytical group will be involved. Specific Tipster products are TBD.



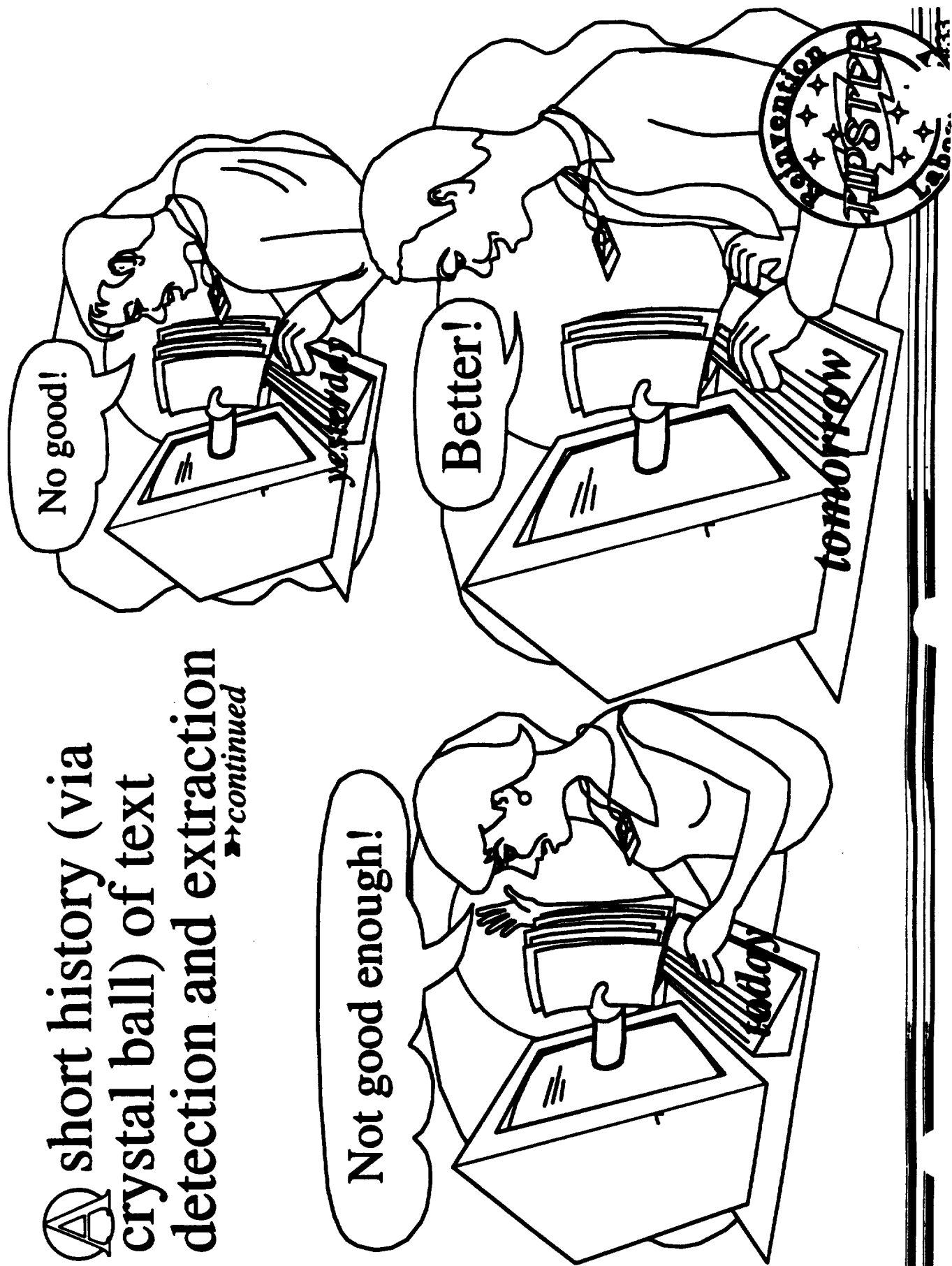
Reinvention Lab Rationale

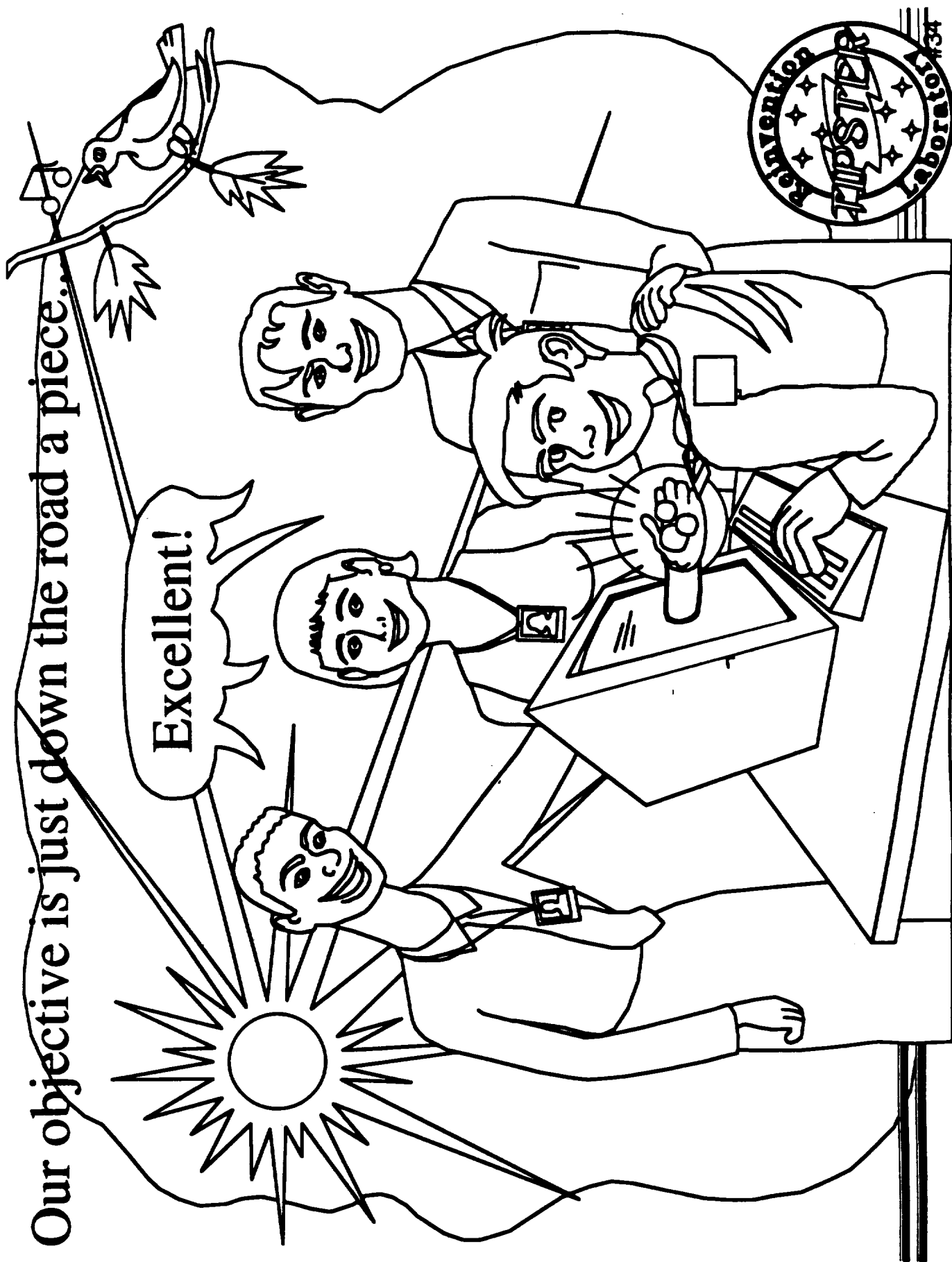
- Tipster has customer focus
- Tipster has broken interagency barriers
- Tipster is teamwork
- Greater visibility will leverage Tipster success within and outside the Government



A short history (via crystal ball) of text detection and extraction

→ continued





INQUIRIES?

- TIPSTER: DR. GEORGE DODDINGTON
ARPA/SISTO
3701 N. FAIRFAX DRIVE
ARLINGTON, VA 22203
(703) 696-2259
(703) 696-2202 (FAX)
doddington@arpa.mil
- MUC: MS. BETH SUNDHEIM
NCCOSC/NRAD
NCCOSC RDTE DIV 44208
53140 GATCHELL ROAD
SAN DIEGO, CA 92152-7420
(619) 553-4145
(619) 553-4149 (FAX)
sundheim@nosc.mil
- TREC: DR. DONNA HARMON
NIST
BUILDING 225, A216
GAITHERSBURG, MD 20899
(301) 975-3569
(301) 975-2128 (FAX)
harman@magi.ncsi.nist.gov

8

CorpTech

1. Corporate Technology Information Services, Inc. (CorpTech)

- ◆ Directory publisher Woburn, MA
 - classifies high technology companies
 - publishes *Corporate Technology Directory*, and *Corporate Technology Database*
 - 10 years ago created classification system because none existed (the classification system is continuously modified to keep up with changes in technology)
 - currently track 40,000 companies with 3,000 product codes (concepts)

2. CorpTech codes ("concepts")

- ◆ Interest has increased in the codes in the last few years as they have improved and become more widely known.
 - SBA RFP 7/6/94 "Select a definition of 'high tech' ...which will be the *Corporate Technology Directory*, or a superior source."
 - University of California (Patents & Trademarks) licenses codes to classify all patent work.
 - Iowa State University approaches companies with certain codes to find technology licensing candidates.

3. Indexing with CorpTech codes ("concepts")

- ◆ Need to find information with a controlled vocabulary
- ◆ Need to correlate information between many databases (full text & fielded)
- ◆ Need to find information with many synonyms or different representations of the same thing...without expert knowledge (EXAMPLE CorpTech Vol 1)

4. Automating The Retrieval of Technology Concepts with software

Or: How to use the CorpTech Codes (15,000 "technology terms" mapped into 3,000 "concepts" organized hierarchically into 17 industries) to facilitate searching full text databases.

Approach #1 -- The Expert System

- ◆ Allow the searcher to define the "concept" to be searched for (using key words, or a cascading menu through a hierarchy)
- ◆ Using the command structure of the search engine in use, construct as specific a query as possible from all the "technology terms" known to relate to the selected "concept"

Advantages

- Augments rather than replaces existing search methods
- Requires no special processing of data, so existing massive databases can use method without reloading
- Prototype running today for environmental "concepts" using the Inquiry engine (UMASS, Amherst) and Mosaic over the Internet
- Prototype can readily be extended:
 - ♠ To other technology industries and also non-technology industries
 - ♠ To other search engines
- Allows the synonym list of "technology terms" to be continually updated to reflect advances in technology, which, in turn, will allow the retrieval of documents using the latest "technology terms" even if they are unknown to the searcher

Disadvantages

- Terms that have different meanings based on context cannot be selected based on any one specific meaning.
- There is no feedback loop to automatically add new "concepts" and "technology terms"
- Limited by the sophistication of the search engine in use
- Requires modification of the search engine user interface

A Mosaic based demo of the expert system is available over the Internet at:

<http://cjit/corptech/homepage.html>

CITR, CS, UMASS, EDU

Corporate Technology Information Services, Inc. 12 Alfred Street Woburn, MA 01801
(800) 333-8036 (617) 932-3939 FAX (617) 932-6335

Approach #2 -- The Pre-Tagging System

- ◆ Tagging software analyses full text documents, appending "tags" relating to each "concept" identified, and a confidence level for each
- ◆ A profile is built of each document defining, based on tag density, the primary focus to also be tagged
- ◆ After tagging, phrases identified by the system as relating to technology or product, can be saved and fed back to CorpTech for editing and potential addition to the authority files as new "concepts" or "technology terms"
- ◆ The searcher "sees" the same interface, selecting "concepts" in the same manner, but now can further refine the search. For example by requesting only documents where the selected "concept" is within the primary focus of the document.

Advantages

- Augments rather than replaces existing search methods
- More precise searching than the "Expert System" approach, with fewer false hits:
 - ♠ System might only tag the phrases "visual display", "VDT", or "monitor" with the concept for "Computer CRT" if the document (or adjacent paragraphs) relate to computers.
 - ♠ Wildcards in "technology terms" can be substituted for different synonym lists depending on the other topics tagged in the document, reducing the number of unique terms to be used.
- Much simpler implementation of hierarchical searches
- Allows searching across multiple databases in different languages

Disadvantages

- Tagging software is not yet operational
- Databases must be tagged and then reloaded before system can work
- Requires modification of the search engine preprocessor and user interface

Questions to: Andrew Campbell, President

Corporate Technology Information Services, Inc. 12 Alfred Street Woburn, MA 01801
(800) 333-8036 (617) 932-3939 FAX (617) 932-6335

CORPTECH® Major Product Groups



AUT / Factory Automation

AUT-AS Assembly manufacturing systems
AUT-AT Automatic test eqp
AUT-FM Flexible manufacturing systems
AUT-MC Manufacturing control systems
AUT-ME Manufacturing measuring eqp
AUT-MH Materials handling/shipping eqp
AUT-MT Machine tools/presses/eqp
AUT-MV Machine vision systems
AUT-RO Robots/robotic eqp
AUT-SV Automation services
AUT-ZZ Other automation



BIO / Biotechnology

BIO-AN Animal biotechnology systems
BIO-CC Cell culture technologies
BIO-CT Catalysts
BIO-EL Bioprocess eqp
BIO-EO Biotechnology eqp
BIO-GE Genetic engineering systems
BIO-IM Immunology systems
BIO-MS Biomass/biochemicals
BIO-MT Biomaterials
BIO-NZ Enzyme technology systems
BIO-PE Proteins/protein engineering sys
BIO-PL Plant biotechnology systems
BIO-SV Biotechnology services
BIO-ZZ Other biotechnology



CHE / Chemicals

CHE-AP Application specific chemicals
CHE-DY Chemical dyes/pigments
CHE-EX Explosive chemicals
CHE-IG Industrial gases
CHE-IN Inorganic chemicals
CHE-OR Organic chemicals
CHE-PT Petrochemicals/petroleum
CHE-RE Reagents
CHE-SV Chemical services
CHE-ZZ Other chemicals



COM / Computer Hardware

COM-AI Artificial intelligence
COM-AX Accessories/components
COM-BU Business equipment
COM-CB Computer boards
COM-CM Computer memory systems
COM-CN Converters
COM-CP CPUs
COM-IN Computer input devices
COM-MC Microcomputers
COM-MF Mainframes
COM-MN Minicomputers
COM-MO Monitors
COM-MS Supercomputers
COM-OU Computer output devices
COM-PC Peripheral controllers
COM-SN Special needs hardware
COM-SP Specialized computers
COM-SV Computer services
COM-TR Terminals
COM-ZZ Other computers



DEF / Defense

DEF-AS Anti-airborne warfare eqp
DEF-CC Command/control eqp
DEF-EW Electronic warfare eqp
DEF-GR Defense ground support eqp
DEF-MI Missiles and related eqp
DEF-OR Ordnance systems/eqp
DEF-ST Surveillance/tracking sys
DEF-SV Defense/government services
DEF-TS Training/simulation sys
DEF-WA Defense warning sys
DEF-ZZ Other defense/military



ENR / Energy

ENR-AL Alternative energy systems
ENR-CN Conventional energy eqp
ENR-CO Energy conservation eqp
ENR-EM Energy management systems
ENR-EP Electrical power sources
ENR-FU Fuels
ENR-HV HVAC equipment
ENR-LI Lighting systems
ENR-ME Meteorological eqp
ENR-MI Mining equipment
ENR-OG Oil/gas recovery eqp
ENR-ST Energy storage eqp
ENR-SV Energy services
ENR-TB Turbines
ENR-TR Electrical power transmission
ENR-TS Training/simulation systems
ENR-ZZ Other energy eqp



ENV / Environmental

ENV-AN Environmental analysis eqp
ENV-CP Control/prevention eqp
ENV-SV Environmental services
ENV-TR Treatment/abatement eqp
ENV-WD Waste disposal eqp
ENV-ZZ Other environmental



MAN / Manufacturing Eqp.

MAN-AE Aerospace manufacturing eqp
MAN-BP Biotech/pharmaceutical mfg eqp
MAN-CP Chemical processing eqp
MAN-CO Computer manufacturing eqp
MAN-EL Electrical manufacturing eqp
MAN-EP Electronic/photonics mfg eqp
MAN-MA Materials processing eqp
MAN-SV Manufacturing related services
MAN-ZZ Other manufacturing eqp



MAT / Advanced Materials

MAT-AB Abrasives
MAT-AD Adhesives
MAT-AM Additives/modifiers
MAT-CE Ceramics and related materials
MAT-CM Composites
MAT-CN Construction materials
MAT-CO Coatings/coating materials
MAT-DE Desiccants
MAT-EL Electronic materials
MAT-FO Foams/foamed materials
MAT-FR Fillers/reinforcements
MAT-ME Metals and alloys
MAT-MO Monomers
MAT-NF Nonmetallic minerals
MAT-NU Nuclear materials
MAT-OL Oils/lubricants
MAT-PO Polymers
MAT-SV Advanced materials services
MAT-TX Textiles/fibers
MAT-ZZ Other materials



MED / Medical

MED-DE Dental eqp
MED-DG Medical diagnostic eqp
MED-HA Handicap aids
MED-HH Home health care products
MED-IM Implants/prostheses
MED-MO Medical monitoring eqp
MED-RE Rehabilitation devices
MED-RP Medical reproduction eqp
MED-SU Surgical/medical eqp
MED-SV Medical services
MED-TH Medical therapeutic eqp
MED-VT Veterinary eqp
MED-ZZ Other medical eqp



PHA / Pharmaceuticals

PHA-AH Antihistamine drugs
PHA-AI Anti-infective agents
PHA-AN Antineoplastic agents
PHA-AU Autonomic drugs
PHA-BL Blood products
PHA-CN Central nervous system drugs
PHA-CV Cardiovascular drugs
PHA-DI Diagnostic agents
PHA-EX Expectorants/antitussives
PHA-EY EENT preparations
PHA-GI Gastrointestinal drugs
PHA-HO Hormones/synthetic substitutes
PHA-LA Local anesthetics
PHA-ME Metabolic/water balance agents
PHA-OC Over-the-counter drugs
PHA-PA Pharmaceutical packaging
PHA-PU Pulmonary agents
PHA-RP Reproductive organ drugs
PHA-SE Serums and related agents
PHA-SK Skin/mucous membrane drugs
PHA-SV Pharmaceutical services
PHA-VI Vitamins
PHA-VT Veterinary pharmaceuticals
PHA-ZZ Other pharmaceuticals



PHO / Photonics

PHO-AO Acousto-optic eqp
PHO-CA Camera and related eqp
PHO-DI Displays
PHO-FO Fiber optics and related eqp
PHO-LA Lasers/laser related eqp
PHO-OE Optoelectronic devices
PHO-OP Optics and related eqp
PHO-SV Photonics services
PHO-ZZ Other photonics eqp



SOF / Computer Software

SOF-AC Accounting software
SOF-AI Artificial intelligence software
SOF-BN Banking software
SOF-CN Construction software
SOF-CS Communications systems software
SOF-DM Database/file management sof
SOF-ED Educational/training software
SOF-FM Facilities management software
SOF-FN Financial analysis/management
SOF-GO Government software
SOF-HL Health services software
SOF-IN Insurance software
SOF-LE Legal software
SOF-LI Library software
SOF-MA Manufacturing software
SOF-ME Media communications software
SOF-NP Non-profit organization sof
SOF-NR Natural resource mgmt software
SOF-OA Office automation software
SOF-PD Program development software
SOF-PM Project management software
SOF-PU Public utilities software
SOF-RE Real estate software
SOF-SM Sales/marketing software
SOF-SR Service industry software
SOF-SV Software services
SOF-TS Transportation software
SOF-TR Technical/scientific software
SOF-UT Unity systems software
SOF-WD Warehousing/distribution sof
SOF-ZA Applications software nec
SOF-ZZ Other software



SUB / Subassemblies & Components

SUB-CE Electronic connectors
SUB-CL Electrical connectors
SUB-CM Mechanical connectors
SUB-EM Electromechanical devices
SUB-ES Electronic subsystems
SUB-ET Electron tubes
SUB-ME Mechanical devices
SUB-PC Passive components
SUB-SE Semiconductors/devices
SUB-SV Electronics services
SUB-TR Transducers
SUB-ZZ Other subassemblies



TAM / Test & Measurement

TAM-AN Analyzers
TAM-CA Calibrators
TAM-CR Counters/recorders
TAM-DE Detection eqp
TAM-EM Measurement eqp
TAM-OS Oscilloscopes
TAM-PV Process variable controllers
TAM-SC Scientific/laboratory eqp
TAM-SS Security/safety eqp
TAM-SV Test & measurement services
TAM-ZZ Other test & measurement eqp



TEL / Telecommunications

TEL-AV Audio/video eqp
TEL-BR Broadcasting/receiving eqp
TEL-CI Communications interfaces
TEL-CS Communications security devices
TEL-CC Data concentration eqp
TEL-EM Electronic mail eqp
TEL-MX Multiplexers/modems
TEL-NW Networks and components
TEL-SI Signal-related eqp
TEL-SM Satellite/microwave eqp
TEL-SV Telecommunications services
TEL-TD Telecom distribution eqp
TEL-TE Telephone/voics eqp
TEL-TR Transmission systems/eqp
TEL-ZD Data communications eqp nec
TEL-ZZ Other telecommunications eqp



TRN / Transportation

TRN-AI Aircraft
TRN-AS Major aircraft systems
TRN-GS Ground support eqp
TRN-GT Ground transportation eqp
TRN-IP Marine systems/eqp
TRN-PR Air/spacecraft propulsion eqp
TRN-SI Signaling eqp
TRN-SR Spacecraft/space vehicles
TRN-SS Space systems
TRN-SV Transportation services
TRN-TS Traffic control eqp
TRN-TS Training/simulation eqp
TRN-ZZ Other transportation eqp



ZZZ / Holding Companies

ZZZ-HC Parent/holding company
ZZZ-HC Subsidiary/holding company
ZZZ-ZZ Other high technology nec

Applied Genetics Laboratories, Inc. (Continued)

Execs: Dr. Maria H. Lugo, Ph.D., Associate Director (Sales, Purchasing, Personnel)

CorpTech	SIC	Product Description	Govt: Prime
810-SV-C	8071	Cell line characterization services	
810-SV-LA	8071	Pharmaceutical testing services	
810-SV-R	8731	Gene mapping services	
810-SV-Z	8071	Cytogenetics services	

Applied Geotechnology Inc.

PMW: Kelly/Strazer Associates, Inc.

PO Box 3885
Bellevue, WA 98009
(206) 453-8383

Fax: (206) 646-9523

Owned: Private
Formed: 1977
T/MEH/33269

Provider of geotechnical engineering services. The company provides investigation and remediation services of hazardous waste in soil and ground water, soil analysis, water analysis, pollution control consulting and waste recovery services. Services are provided to multiple industries.

Sales: \$10 million; no international sales
Empl: 115 employees; 15% employment growth in prior year
Execs: John Newby, President (C.E.O.)

Mackey Smith, Vice President/Chief Hydrogeologist (R&D)
Donald E. Bruggers, PE, Director of Geotech Engineering (Engineering)
Ms. Dani Schalka, Administrative Manager (Finance, Administration)

CorpTech	SIC	Product Description
ENV-SV-AS	8734	Soil analysis services
ENV-SV-AM	8734	Water analysis services
ENV-SV-CP	8999	Pollution control consulting services
ENV-SV-WR	4959	Waste recovery services
ENV-SV-WT	4959	Hazardous waste remediation services

Applied Graphics, Inc.

1960 Shelburne Rd.
Shelburne, VT 05482
(802) 985-3341

Fax: (802) 985-8302

Owned: Private
Formed: 1957
T/DMP/12487

Provider of computer systems consulting, design and implementation services. Services are provided to multiple industries.

Sales: \$12 million; no international sales
Empl: 53 employees; 12% employment reduction in prior year
Execs: William Stevens, President (C.E.O., Finance, Sales, Marketing)
Ms. Katherine Stevens, Vice President

CorpTech	SIC	Product Description
COM-SV-CC	7373	Computerization consulting services

Applied Holographics Corp. (FKA)

See: CFC Applied Holographics

Applied Holographics PLC

England

Parent/holding company with high-tech operating units involved in the advanced material and photonics industries.

U.S. high-tech operating units:
CFC Applied Holographics, IL

Applied Hydro-Acoustics Research, Inc.

AKA: AHA

656 Quince Orchard Rd., Suite 710
Gaithersburg, MD 20878
(301) 840-9722

Fax: (301) 975-1067

Owned: Private
Formed: 1973
T/MEH/43285

Provider of research and development services specializing in the technologies of submarine warfare, underwater acoustics, sonar, ship and truck noise silencing. Services are primarily provided to the U.S. Navy.

Sales: Under \$1m; no international sales
Empl: 17 employees; 6% employment growth in prior year
Execs: Dr. John McNicholas, Ph.D., President (C.E.O., Finance, R&D, Sales)
Francis Ryder, Vice President (Manufacturing)

CorpTech	SIC	Product Description	Govt: Prime
DEF-SV	8711	Submarine warfare research and development services	
TEL-SV-R	8711	Underwater acoustics research and development services	
TRN-SV-R	8711	Submarine warfare research and development services	
TRN-SV-R	8711	Sonar technology research and development	
TRN-SV-R	8711	Ship silencing research and development services	
TRN-SV-R	8711	Truck noise research and development services	
TRN-SV-V	8711	Ship silencing research and development services	

Applied I (FKA)

See: Tutsm Products

Applied Image, Inc.

1653 East Main St.
Rochester, NY 14609
(716) 482-0300

Fax: (716) 288-5989

Owned: Private
Formed: 1979
W/JMH/11794

Manufacturer of micromaging and precision imaging products. Products include encoders, micromasks, reticles and density wedges. Products are used to measure density of light and characteristics of film. Micromasks are microscopic circuitry used in integrated circuits and hybrid circuits. Products are sold to the electronics industry.

Sales: \$1m to \$2.5m; 2.5% to 10% international
Empl: 25 employees; no employment change in prior year
Execs: Bruno B. Glavich, CEO/President (C.E.O., Finance, R&D, Tech. Transfer)
Eduardo J. Navarro, VP of Marketing (Sales, Marketing, Int'l)
Allen Bolognesi, Production Manager (Manufacturing)
Gary Rief, Engineering Manager (Engineering)
Ms. Shari Lamphorn, Purchasing Agent (Purchasing)

Applied Image, Inc. (Continued)

CorpTech	SIC	Product Description
AUT-MC-D	3525	Encoders
MAN-EP-M	3559	Micromasks
PHO-OP-PQ	3827	Reticles
TAM-ME-P	3820	Density wedges

Applied Image & Information

Unit of Day & Zimmermann, Inc., PA

280 King of Prussia Rd.
Radnor, PA 19087
(610) 976-8883

This company has been identified by CorpTech and is involved in the software industry. At publication time the profile had not been completed.

Applied Imaging Corp.

2340A Walsh Ave., Bldg. F

Santa Clara, CA 95051

(408) 562-0250 (800) 634-3622 Fax: (408) 562-0264

Owned: Private
Formed: 1987
T/JFD/47751

Manufacturer of medical computer systems. Genevision™ and Cytoscan™ are utilized for cytogenetics and SpeedReader™ is utilized for DNA sequencing. Products are sold to the medical industry.

Sales: \$8.5 million (estimate)
Empl: 52 employees; no employment change in prior year
Execs: Abraham Coriat, President/CEO (C.E.O., Sales, R&D)
Neil Woodruff, Chief Operating Officer (Finance)
Steve Rudy, VP of Marketing (Marketing)

CorpTech	SIC	Product Description
COM-SP-IM	3571	Medical computer systems - Genevision™
COM-SP-IM	3571	Medical computer systems - Cytoscan™
COM-SP-IM	3571	Medical computer systems - SpeedReader™

Applied Immune Sciences, Inc.

5301 Patrick Henry Dr.

Santa Clara, CA 95054

(408) 492-9200 (800) 247-8881 Fax: (408) 980-5888

Owned: Public (AISX)
Formed: 1984
W/HSM/44838

Developer of immune system therapies and technologies. The products isolate specific cells or protein from the patient's blood which can then be eliminated, modified, or numerically expanded for reinfusion into the patient's bloodstream. Products include the CELLector™, which can be used in AIDS treatment, bone marrow transplantation and cancer and leukemia treatment; and the Cytasorb™, which is used in plasma therapy. A laboratory research version is sold under the MicroCELLector™ name. Products are sold to hospitals, clinics and scientific research institutions. This company was capitalized by private investment, venture capital and corporate investment.

Sales: \$1.5 million; over 25% international
Empl: 184 employees; 64% employment growth in prior year
Execs: Dr. Thomas Okarma, President/CEO (C.E.O.)
James Smith, Senior Vice President/COO (Administration)
Craig Huffaker, Senior Vice President/CFO (Finance)
Jack Anthony, Vice President (Sales, Marketing, Int'l)
David Okronghy, VP of Advanced Applications Development (R&D, Engineering)
Ms. Jean LaDouceur, VP of Operations (Manufacturing)

CorpTech	SIC	Product Description
MED-TH-S	3841	Cell therapy systems - CELLector™
MED-TH-S	3841	Plasma therapy systems - Cytasorb™

Applied Industrial Materials Corp.

AKA: AIMCOR

1 Pkwy. North, Suite 400

Deerfield, IL 60015

(708) 940-8700 Fax: (708) 940-0811

Owned: Private
Formed: 1986
T/MEH/45749

Parent/holding company with high-tech operating units involved in the energy and advanced material industries.

Sales: \$66 million (estimate); over 25% international
Empl: 450 employees; no employment change in prior year
Execs: Charles P. Gallagher, Chairman of the Board (C.O.B., R&D)
Wayne Kocourek, President (C.E.O., Sales)
Jon M. Burns, VP of Human Resources and Administration (Personnel)
James Ehren, Chief Financial Officer (Finance)

CorpTech	SIC	Product Description
ZZZ-IC	6719	Parent/Holding company

U.S. high-tech operating units:
Applied Industrial Materials Corp. / Metals and Minerals Division, PA

Applied Industrial Materials Corp. /

Metals and Minerals Division

AKA: AIMCOR / Metals and Minerals Division;

Unit of Applied Industrial Materials Corp., IL

Park Ridge 1 Office Ctr., Suite 200, Commerce Dr.

Pittsburgh, PA 15275

(412) 788-1860 (800) 541-8594 Fax: (412) 788-1822

Owned: Private unit
Formed: 1986
T/SRP/45623

Producer of minerals, metals and alloys. Products include ferrichrome, magnesium, olivine, vanadium and fluorapatite. Products are sold to the steel, aluminum and chemical industries.

Sales: \$9.0 million (estimate); 2.5% to 10% international
Empl: 61 employees; 7% employment reduction in prior year
Execs: Charles W. Kopeck, President

CorpTech	SIC	Product Description
MAT-ME-FV	3320	Ferrichrome
MAT-ME-G	3339	Olivine
MAT-ME-G	3339	Magnesium
MAT-ME-Z	3399	Vanadium

* - company listed for the first time. AKA: also known as. FKA: formerly known as. DBA: doing business as. PMW: previously merged with. SHC: subsidiary holding company.
AUT: automation BIO: biotechnology. CHE: chemicals. COM: computers. DEF: defense/military. ENR: energy. ENV: environmental. MAN: manufacturing. MAT: advanced materials. MED: medical.
PHA: pharmaceuticals. PHO: photonics. SCF: software. SUB: subassemblies/components. TAM: test & measurement. TEL: telecommunications. TRN: transportation. ZZZ: holding companies.
Corporate Technology Directory - © Copyright 1994 Corporate Technology Information Services, Inc. Tel: (617) 932-3939 - All rights reserved.

Applied Industrial Materials Corp. / Metals and Minerals Division (Continued)
MAT-WF 3299 Fluorspar

Applied Informatics, Inc.

15104 Saint Thomas Church Rd.
Upper Marlboro, MD 20772
(301) 627-6650

Owned: Private
Formed: 1984
T/CJD/23202

Provider of custom applications software programming services. Services are provided to multiple industries. This company was capitalized by private investment.

Sales: \$3.92 million (estimate); no international sales
Empl: 8 employees; 14% employment growth in prior year
Execs: Samuel S. Cottrell, President (C.E.O., Manufacturing, R&D, Personnel)
Stephen J. Kelly, Sales Manager (Sales)

CorpTech SIC Product Description
SOF-SV-PA 7371 Custom applications software development services

Applied Information Development, Inc. (PMW)

See: Computer Power Group, Ltd.

Applied Information for Marketing, Inc.

AKA: AIM

15 Katchum St.
Westport, CT 06880
(203) 226-0318

Fax: (203) 227-8969

Owned: Private
Formed: 1988
W/EJS/39892

Developer of market analysis software which analyzes trade expenditures. Software runs on PC-based systems. Products are sold to multiple industries.

Sales: \$3.1 million (estimate); no international sales
Empl: 27 employees; 237% employment growth in prior year
Execs: Joseph Anstey, President (C.E.O., Sales, R&D, Finance)
William Arndt, Systems Manager (Manufacturing, M.I.S.)

CorpTech SIC Product Description
SOF-SH-A 7372 Market analysis software - PROMOMAX™

Applied Information Management Sciences, Inc.

AKA: AIMS, Inc.

PO Box 7680
Monroe, LA 71211
(318) 323-2467

(800) 729-2467 Fax: (318) 322-3472

Owned: Private
Formed: 1968
T/CLM/12889

Developer of accounting software. One integrated program, COMPAS™, is a complete oil marketer's peripheral accounting system. The software includes inventory control, accounts payable, accounts receivable, payroll and personnel accounting, order entry and general ledger. AIMS™ is an integrated package and includes inventory control, accounts payable, accounts receivable, general ledger and sales order entry used mainly for wholesale/retail distribution; timing and billing are included as options. The software runs on IBM PCs and compatibles. Products are sold to wholesale distributors. This company was capitalized by private investment.

Sales: \$1.0 million; 0% to 2.5% international
Empl: 15 employees; 25% employment growth in prior year
Execs: Dr. Bob Canterbury, President (C.E.O., Sales)
Robert Canterbury, Vice President (Marketing, Personnel)
Ms. Gail Babcock, Director of Technical Services (R&D)
John Schrum, Director of Purchasing (Purchasing)
Stan Hargus, Director of Marketing
Ms. Joann Lewis, Financial Manager (Finance)

CorpTech SIC Product Description
SOF-AC-B 7372 Billing software
SOF-AC-G 7372 General ledger software
SOF-AC-P 7372 Accounts payable software
SOF-AC-R 7372 Accounts receivable software
SOF-AC-Y 7372 Payroll and personnel accounting software
SOF-SH-D 7372 Oil marketing software package - COMPAS™
SOF-MD-I 7372 Wholesale/retail store inventory control software package - AIMS™
SOF-MD-O 7372 Order entry software

Applied Information Solutions, Inc.

AKA: AIS

1600 Wynton St., Suite 300
Denver, CO 80202
(303) 893-8936

Fax: (303) 893-5046

Owned: Private female
Formed: 1985
T/DJC/45244

Developer of partnership administration software, INGOT™, that runs on IBM personal computers and compatibles. Products are sold to the financial, real estate and accounting industries. This company was capitalized by private investment.

Sales: \$1.9 million (estimate); 0% to 2.5% international
Empl: 17 employees; no employment change in prior year
Execs: Ms. Ann C. Bennett, President/CEO (C.E.O.)
Bob Grizzle, Vice President (Marketing)
Russell York, Vice President (Finance)
D.J. Mattoon, Technical Director (R&D, M.I.S.)
Ms. Deborah Spencer, Corporate Development Advisor (Corporate Dev., Sales)

CorpTech SIC Product Description
SOF-FN-I 7372 Fund/Partnership administration software - INGOT™

Applied Information Systems, Inc.

AKA: AIS

100 Europa Dr., Suite 555
Chapel Hill, NC 27514
(919) 942-7801

(800) 334-5510

Tlx: 701 365 APPLIED INFO-UD

Owned: Private
Formed: 1972
T/TAV/15057

Developer of data entry utilities software, engineering workstation software and file transfer software, a DEC/Unsys communications system. Products are sold to multiple industries.

Applied Information Systems, Inc. (Continued)

Sales: \$1.0 million; over 25% international
Empl: 10 employees; no employment change in prior year
Execs: Arthur W. Coston, President (C.E.O., Finance)
Ms. Anne Hartley, Vice President (Sales)
Goutham Penmetsa, Software Manager (R&D)

CorpTech SIC Product Description
SOF-CS-M 7372 File transfer software - BURCOM™
SOF-DA-E 7372 Data entry utilities software - Easy-Entry™
SOF-DA-S 7372 Engineering spreadsheet workstation software - XESS™
SOF-TS-EG 7372 Engineering spreadsheet workstation software - XESS™

Govt: Sub

Applied Information Systems, Inc.

1640 Powers Ferry Rd., Bldg. 8
Marietta, GA 30067
(404) 953-1903

Owned: Private
Formed: 1977
T/SRP/15493

Developer of transportation and accounting software and provider of consulting and custom programming services. The transportation software is used by the trucking industry and includes general ledger, fuel and mileage tracking, dispatching, freight billing, driver settlements, driver logs, accounts payable/receivable and payroll functions. The integrated accounting software has general ledger, accounts payable/receivable, invoicing/billing and payroll capacities. The software runs on IBM PC and compatible computers. The company also provides custom programming services and systems analysis and development consulting services. Products and services are sold to multiple industries.

Sales: \$5.8 million (estimate); no international sales
Empl: 5 employees; no employment change in prior year
Execs: Timothy C. Ponder, President

CorpTech SIC Product Description
COM-SV-CA 7379 Systems analysis services
COM-SV-CC 7373 Computerization consulting services
SOF-AC-B 7372 Invoicing/billing software
SOF-AC-G 7372 General ledger software
SOF-AC-I 7372 Integrated accounting software
SOF-AC-P 7372 Accounts payable software
SOF-AC-R 7372 Accounts receivable software
SOF-AC-Y 7372 Payroll and personnel accounting software
SOF-SV-PA 7371 Applications software custom programming services
SOF-SV-PS 7371 Programming aids software custom programming services
SOF-SV-PS 7371 Productivity software custom programming services
SOF-SV-PS 7371 Utility software custom programming services
SOF-SV-PS 7371 System software custom programming services
SOF-MD-D 7372 Trucking software
SOF-MD-FT 7372 Truck/fleet management software

Applied Innovations Software, Inc. (AKA)

See: AI Software, Inc.

Applied Instruments Co.

11 Roosevelt Ave.
Valley Stream, NY 11581
(516) 872-8111

Fax: (516) 872-8112

Owned: Private
Formed: 1948
T/CLM/11630

Parent/holding company with high-tech operating units in the photonics industry.

Sales: Under \$1m; 2.5% to 10% international
Empl: 12 employees; no employment change in prior year
Execs: Albert Joseph, President

CorpTech SIC Product Description
ZZI-WC 6719 Parent/Holding company
U.S. high-tech operating units:
Applied Instruments Co. / Optical Division, NY

Applied Instruments Co. / Optical Division

Unit of Applied Instruments Co., NY

11 Roosevelt Ave.
Valley Stream, NY 11581
(516) 872-8111

Fax: (516) 872-8112

Owned: Private unit
Formed: 1984
W/CLM/20922

Manufacturer of optical components including heat absorbing and resisting filters, spherical and aspheric fire polished lenses, partial reflectors, prisms and windows. Lenses are projection lenses used to condense or project light. Products are used in many applications, including as components in projection equipment.

Sales: Under \$1m; 2.5% to 10% international
Empl: Under 10 employees
Execs: Albert Joseph, President

CorpTech SIC Product Description
PHO-OP-FA 3827 Heat absorbing filters
PHO-OP-FI 3827 Heat resisting filters
PHO-OP-LD 3827 Aspheric fire polished lenses
PHO-OP-LZ 3827 Spherical lenses
PHO-OP-PP 3827 Prisms
PHO-OP-PR 3827 Partial reflectors
PHO-OP-PR 3827 Reflectors in glass and aluminum
PHO-OP-PW 3827 Windows

Govt: Prime

Applied Intelligent Systems, Inc.

AKA: AISI; PMW; VSP, Inc.

110 Parkland Plaza
Ann Arbor, MI 48103
(313) 995-2035

Fax: (313) 995-2138

Owned: Private
Formed: 1976
W/MEH/11054

Manufacturer of machine vision computers and software, used for inspection, guidance, control and identification. Products are used in the automotive and pharmaceutical industries and by manufacturers of semiconductor and electronic packaging and assemblies. The software runs on a parallel processor which is part of the machine vision computer. Products are sold to the semiconductor and electronics industries. This company was capitalized by private investment, venture capital and corporate investment.

* - company listed for the first time; AKA: also known as; FKA: formerly known as; DBA: doing business as; PMW: privately merged with; SIC: subsidiary holding company; AUT: automation; BIO: biotechnology; CHE: chemicals; COM: computers; DEF: defense/military; ENR: energy; ENV: environmental; MAN: manufacturing; MAT: advanced materials; MED: medical; PHA: pharmaceuticals; PHO: photonics; SOF: software; SUB: subassemblies/components; T&M: test & measurement; TEL: telecommunications; TRN: transportation; ZZZ: holding companies; Corporate Technology Directory - © Copyright 1994 Corporate Technology Information Services, Inc. Tel: (617) 932-3939 - All rights reserved.

COMPAT

Business Indexes

Company Name Index

Company	St	Phone	Sales	Own	V-Ps
Sterilization Technical Services, Inc. (FKA)	NY	(716) 533-1672	3	PR	4-701
Steris Corp.	OH	(216) 354-2600	4	PU	4-678
Steris Laboratories	AZ	(602) 278-1400	7	PRu	4-678
* Steri Technologies, Inc.	NY	(516) 563-8300	1	PR	4-679
Sterling Detroit Co.	MI	(313) 366-3500	2	PR	4-679
Sterling Drug Inc. (FKA)	NY	(212) 907-2000	9	PUu	4-681
Sterling Electric, Inc.	CA	(714) 474-0520	4	PR	4-679
Sterling Electronics Corporation	TX	(713) 627-9800	7	PU	4-679
Sterling Engineered Products, Inc. (FKA)	MI	(313) 286-0800	8	PUu	2-99
* Sterling Engineering Corp.	CT	(203) 379-3366	-	PR	4-679
Sterling Glass, Inc. (AKA)	IL	(708) 864-6900	1	PR	4-680
Sterling Group (PMW)	MO	(314) 621-5700	5	PR	4-121
Sterling Health	NY	(212) 907-2000	9	PUu	4-679
Sterling, Inc.	MI	(414) 354-0970	5	PR	4-679
Sterling Inc. / Steritech Division	MI	(414) 354-0970	3	PRu	4-679
Sterling Industries PLC	England			FO	4-679
Sterling Laboratories (FKA)	TX	(913) 926-0000	0	PR	4-680

Product Indexes

Technology Index

Technology/Product	CorpTech Code	SIC Code	Who Mkt	Why
Archiving software, file	SOF-UT-J	7372	1-1162	
Arc jet engines	TRN-PR-RR	3764	1-1329	
Arc lamp power supplies	PHO-DE-LY	3641	1-1024	
Arc lamps	PHO-DE-LA	3641	1-1023	
Arc/spark emission spectrometers	TAM-SC-ISA	3826	1-1276	
Arc welding electrodes	AUT-AS-W	3548	1-754	
Area navigation eqp. aircraft	TRN-AS-EN	3812	1-1322	
Argon, industrial gases	CHE-IG	2813	1-809	
Argon-ion lasers	PHO-LA-LGA	3699	1-1020	
Argon lasers	PHO-LA-LGA	3699	1-1020	
Arithmetic logic unit (ALU) ICs	SUB-SE-1K	3674	1-1236	
Armament systems/armament	DEF-OR-K	3483	1-867	
Arming systems	DEF-OR-D	3483	1-866	
Armored vehicles	DEF-GR-V	3711	1-866	
Armor, ordnance eqp	DEF-OR-Z	3489	1-868	
Armor piercing bombs	DEF-OR-B	3483	1-866	

Geographic Index — California

Company	St	Phone	Sales	Own	V-Ps
South El Monte (Cont)					
Equipment Manufacturing Corp.	CA	(818) 575-1644	0	PR	3-143
Lee Pharmaceuticals	CA	(818) 442-3141	4	PU	3-874
Plastic Engineered Components, Inc. / Los Angeles Die Mold Division	CA	(818) 575-3551	4	PRu	4-188
Vacuo Industries	CA	(818) 443-7121	9	PUu	4-997
South Pasadena, CA (ZIP: 910xx)					
Keck-Craig, Inc.	CA	(818) 799-1159	1	PR	3-778
Shirley Software Systems	CA	(818) 441-5121	0	PR	4-538
* Simgraphics Engineering Corp.	CA	(213) 255-0900	-	PR	4-560
South San Francisco, CA (ZIP: 940xx)					
Athena Neurosciences, Inc.	CA	(415) 877-0900	0	PU	2-353
Berlex Biosciences	CA	(415) 952-7070	4	FOu	2-468
Caltag Laboratories, Inc.	CA	(415) 873-6106	1	PU	2-606
* COR Therapeutics, Inc.	CA	(415) 244-6800	1	PU	2-888
Data Control, Ltd.	CA	(415) 873-6791	1	FOu	2-979
Genemed Biotechnologies, Inc.	CA	(415) 952-8193	0	PR	3-307
Genentech, Inc.	CA	(415) 266-1000	9	PU	3-308
Genomix Corp.	CA	(415) 266-1157	-	PR	3-329
Guy F. Atkinson Co. of California	CA	(415) 876-1000	9	PU	3-402
HSQ Technology	CA	(415) 952-4310	4	PR	3-518
Imatron, Inc.	CA	(415) 583-9964	4	PU	3-571
Monster Cable Products, Inc.	CA	(415) 871-6000	5	PR	3-1187
Phaser Systems, Inc.	CA	(415) 952-6300	2	PR	4-155
Polywell Computers, Inc.	CA	(415) 583-7222	3	PR	4-204
* Precision Plastics	CA	(415) 588-4450	1	PRu	4-228
Siva International, Inc.	CA	(415) 589-9600	3	PR	4-567
Swart Industries Corp.	CA	(415) 588-4450	2	PR	4-725
Swart Interconnect	CA	(415) 588-8651	2	PRu	4-725
TOA Electronics, Inc.	CA	(415) 588-2538	2	FOu	4-879
Toshiba America MRI	CA	(415) 872-2722	0	FOu	4-887
Zymed Laboratories, Inc.	CA	(415) 871-4494	3	PR	4-1172
Spring Valley, CA (ZIP: 920xx)					
Advan Retro Technology, Inc.	CA	(619) 670-5100	0	PR	2-83

Who Makes What Index — Photonics

Company	St	Phone	Sales	Own	V-Ps
PHO-FO-PCR — Fiber optic repeaters (SIC: 3827)					
E.I. du Pont De Nemours and Company / Electro-Optics Products	NC	(919) 481-5100	3	PUu	3-47
Honeywell Inc. / Systems and Research Center	MI	(612) 951-7226	9	PUu	3-504
McLAN Technology, Inc.	CA	(408) 752-2770	2	PR	3-1145
Montex USA, Inc.	CA	(619) 453-7905	1	FOu	3-1192
Philips Broad Band Network, Inc.	NY	(315) 682-9105	6	FOu	4-158
* TCL, Inc.	CA	(510) 657-3800	3	PR	4-779
Telco Systems, Inc. / Fiber Optics Division	MA	(617) 551-0300	7	PUu	4-805
PHO-FO-PCS — Fiber optic switches (SIC: 3827)					
Advanced Optical Systems, Inc.	MD	(301) 984-3945	1	PR	2-79
AMP Incorporated / Kaptron	CA	(415) 493-8008	2	PUu	2-232
Dicon Fiberoptics, Inc.	CA	(510) 528-0427	1	PR	2-1065
Fujitsu America, Inc. / Cellular Mobile Telephone Division	TX	(214) 690-6000	6	FOu	3-277
Honeywell Inc. / Optoelectronics Division	TX	(214) 470-4271	5	PUu	3-503
Honeywell Inc. / Systems and Research Center	MI	(612) 951-7226	9	PUu	3-504
NEC Electronics, Inc.	CA	(415) 960-6000	9	FOu	3-1254
Optivision, Inc.	CA	(916) 756-4429	2	PR	4-49
* UCE, Inc.	CT	(203) 838-7509	1	PR	4-939
PHO-FO-PCT — Fiber optic transmitters (SIC: 3827)					
Aborn Electronics, Inc.	CA	(408) 436-5444	1	PR	2-18
American Laser Systems, Inc.	CA	(805) 967-0423	0	PR	2-204
CR International, Inc.	MD	(301) 595-5350	1	PR	2-908
Digital Microwave Corp.	CA	(408) 943-0777	6	PU	2-1075
DSC Communications Corporation	TX	(214) 519-3000	8	PU	2-1118
Hewlett-Packard Co. / Optical Communications Division	CA	(408) 435-7400	9	PUu	3-472
Honeywell Inc. / Systems and Research Center	MI	(612) 951-7226	9	PUu	3-504

Non-U.S. Parent Company Index

Company	U.S. Units	V-Ps
India		
DCM, Ltd.	1	2-1018
HMT International, Ltd.	1	3-492
Rewdale Precision Tools Private Ltd.	1	4-376
Ireland		
East Coast Software, Ltd.	1	3-7
Elan Corporation PLC	2	3-51
* Etos, Inc.	1	3-159
Jefferson Smurfit Group, Ltd.	1	3-727
Jones Group PLC	1	3-749
Powerscreen International PLC	1	4-214
Israel		
* AITECH Systems Ltd.	1	2-125
Dynalene Ltd.	1	2-380

SIC to CorpTech Code Translation Table

SIC Code	CorpTech Code	Name
3600 Electronic eqp. not computer eqp	PHO-DE-A	Photorelays
	PHO-DE-B	Blackbody sources
	PHO-DE-C	Photocurrent devices
	PHO-DE-D	Photodetectors
	PHO-DE-E	Photoelectric emitters
	PHO-DE-F	Photoelectric interrupters
	PHO-DE-G	Photoelectric isolators
	PHO-DE-H	Laser modulators
	PHO-DE-I	Light modulators
	PHO-DE-J	Other photoelectric modulators
	PHO-DE-K	nec
	PHO-DE-L	Photoelectric relays
	PHO-DE-M	Photomultipliers
	PHO-DE-N	PhotoSCRs
	PHO-DE-O	Photoreistors
	PHO-DE-P	Photo-electric switches

SOF	Computer Software	Software applications, systems, and services	7370
SOF-AC	Accounting software	Accounting software	7372
SOF-AC-B	Invoicing/billing software	Invoicing and billing software	7372
SOF-AC-F	Fixed asset management sof	Fixed asset management software	7372
SOF-AC-G	General ledger software	General ledger software	7372
SOF-AC-I	Integrated accounting sof	Integrated accounting software	7372
SOF-AC-J	Job costing software	Job costing software	7372
SOF-AC-P	Accounts payable software	Accounts payable software	7372
SOF-AC-R	Accounts receivable sof	Accounts receivable software	7372
SOF-AC-T	Tax preparation/reporting	Tax preparation and reporting software	7372
SOF-AC-U	Purchasing software	Purchasing software	7372
SOF-AC-Y	Payroll/personnel acctg	Payroll and personnel accounting software	7372
SOF-AC-Z	Other accounting software	Other accounting software nec	7372
SOF-AI	Artificial intel software	Artificial intelligence software	7372
SOF-AI-A	Voice technology software	Voice technology software	7372
SOF-AI-C	Sof compilers/interpreters	Software to translate high level language to machine language	7372
SOF-AI-CG	LOGO software compilers	Compilers/interpreters for the AI programming language LOGO	7372
SOF-AI-CI	IPL software compilers	Compilers/interpreters for the AI programming language IPL	7372
SOF-AI-CK	KRL software compilers	Compilers/interpreters for the AI programming language KRL	7372
SOF-AI-CL	LISP software compilers	Compilers/interpreters for the AI programming language LISP and versions of the LISP language	7372
SOF-AI-CO	OPS software compilers	Compilers/interpreters for the AI programming language OPS and any version of the OPS language	7372
SOF-AI-CP	PLANNER compilers	Compilers/interpreters for the AI programming language PLANNER	7372
SOF-AI-CR	PROLOG software compilers	Compilers/interpreters for the AI programming language PROLOG and versions of the PROLOG language	7372
SOF-AI-CS	SAIL software compilers	Compilers/interpreters for the AI programming language SAIL	7372
SOF-AI-CZ	Other AI compilers nec	Other AI programming language compilers/interpreters nec	7372
SOF-AI-E	Expert system software	Software to emulate expert ability in specific fields	7372
SOF-AI-EA	Trend analysis software	Expert system software for trend analysis and prediction	7372
SOF-AI-EB	Business planning sys sof	Expert system software to create a skeletal business plan	7372
SOF-AI-EC	Devel/bldg expert sys sof	Expert system software for development of expert system applications software	7372
SOF-AI-ED	Decision making expert sys	Expert system software to analyze and help solve business	7372
SOF-AI-EI	Industry-specific software	Expert system software designed for a specific industry segment or application	7372
SOF-AI-EN	Needs analysis sys sof	Expert system software to analyze need	7372
SOF-AI-EP	Program generation sys sof	Expert system software for generating application software	7372
SOF-AI-EZ	Other expert systems nec	Other expert system software nec	7372
SOF-AI-L	Natural language software	Natural language software	7372
SOF-AI-N	Neural network software	Software that simulates the neural pathways of the brain	7372
SOF-AI-T	Tutor/instructional expert sy	Expert system software to analyze student errors and conducts natural language dialogue	7372
SOF-AI-V	Visual technology software	Visual technology software	7372
SOF-AI-Z	Other AI software nec	Other artificial intelligence software nec	7372
SOF-BA	Banking software	Software for banking and finance	7372
SOF-BA-A	Automatic teller machine	Automatic teller machine software	7372
SOF-BA-B	Brokerage software	Software to track stock exchange information	7372
SOF-BA-C	Credit union mgmt software	Software to manage credit unions	7372
SOF-BA-D	Account activity software	Software dealing with bank account activity	7372
SOF-BA-F	Multicurrency software	Software for international finance corporations working with multiple currencies	7372
SOF-BA-I	Integrated banking sof	Integrated banking software	7372
SOF-BA-J	Commodities management sof	Commodities management software	7372
SOF-BA-L	Loans/mortgages software	Software to manage loans and mortgages	7372
SOF-BA-M	Credit/collection sof	Software to manage credit, collections, credit cards, etc.	7372
SOF-BA-P	Portfolio management sof	Portfolio management software	7372
SOF-BA-Z	Other banking software	Other banking software nec	7372
SOF-CN	Construction software	Software for construction or contractor management	7372
SOF-CS	Communications sys sof	Communications systems software	7372
SOF-CS-C	Communications control sof	Software for controlling communications	7372
SOF-CS-E	Emulation/simulation sof	Emulators/simulators that integrate otherwise incompatible terminals and computers	7372
SOF-CS-F	Fax software	Software for controlling fax communications	7372
SOF-CS-L	Local area network sof	Local area network (LAN) software	7372
SOF-CS-M	Microcom-mainframe sof	Software for communications and file transfer between	7372

		microcomputers and mainframes	7372
SOF-CS-W	WAN software	Wide area network software	7372
SOF-CS-Z	Other communications sof	Other communications software nec	7372
			7372
SOF-DM	Database/file mgmt sof	Database/file management software	7372
SOF-DM-D	Database dictionaries	Database dictionaries	7372
SOF-DM-F	File management software	File management software	7372
SOF-DM-M	Database mgmt sys (DBMS)	Database management systems (DBMS)	7372
SOF-DM-MH	Hierarchical DBMS sof	Software to present data to users in a tree-like structure	7372
SOF-DM-MR	Relational DBMS software	Software to present data to users in the form of tables of rows and columns	7372
			7372
SOF-DM-MZ	Other DBMS software nec	Other DBMS software nec	7372
SOF-DM-Q	Database query language	Database query language software	7372
SOF-DM-Z	Other database/file mgmt	Other database/file management software nec	7372
			7372
SOF-ED	Educational/training sof	Educational and training software	7372
SOF-ED-A	School/student admin	School and student administrative software	7372
SOF-ED-E	University/adult ed sof	Software for university or adult education	7372
SOF-ED-F	For lang instruction sof	Software to teach foreign languages	7372
SOF-ED-G	K-8 educational software	Software for nursery school through 8th grade education	7372
SOF-ED-M	Music instruction software	Music instruction software	7372
SOF-ED-R	Reading improvement sof	Software to improve reading speed and comprehension	7372
SOF-ED-S	Secondary school software	Software for secondary school education	7372
SOF-ED-T	Typing instruction sof	Typing instruction software	7372
SOF-ED-V	Occupational training sof	Software to train someone in a specific occupational skill	7372
SOF-ED-Z	Other educational software	Other educational/training software nec	7372
			7372
SOF-FM	Facilities management sof	Facilities management software	7372
SOF-FM-C	Communications mgmt sof	Software to manage communications in a facility	7372
SOF-FM-E	Energy analysis/mgmt sof	Energy management software	7372
SOF-FM-H	HVAC software	Software to analyze and control heating, ventilation, and air conditioning	7372
			7372
SOF-FM-I	Integrated facilities mgmt	Integrated facilities management software	7372
SOF-FM-L	Lighting analysis/control	Software to analyze and control the lighting in a facility and control it	7372
			7372
SOF-FM-M	Facilities maintenance softwa	Software to track facility maintenance and repair	7372
SOF-FM-S	Building security software	Building security and management software	7372
SOF-FM-Z	Other facilities mgmt nec	Other facilities management software nec	7372
			7372
SOF-FN	Finan analysis/mgmt sof	Financial analysis/management software	7372
SOF-FN-I	Integrated financial sof	Integrated financial management software	7372
SOF-FN-M	Investment management sof	Investment management software	7372
SOF-FN-P	Planning/analysis software	Analysis and planning software	7372
SOF-FN-T	Templates/models software	Templates and models software	7372
SOF-FN-Z	Other financial sof nec	Other financial analysis/management software nec	7372
			7372
SOF-GO	Government software	Software for federal, state and local governments and government associations	7372
SOF-GO-A	Social services software	Software for governmental social services	7372
SOF-GO-F	Federal government sof	Software for the federal government and federal government associations	7372
			7372
SOF-GO-L	Law enforcement software	Software for law enforcement agencies	7372
SOF-GO-M	Military/defense software	Military or defense applications software	7372
SOF-GO-R	Regulatory compliance sof	Software to monitor and enforce federal government regulations and statutes	7372
			7372
SOF-GO-S	State/local government sof	Software for local and state governments	7372
SOF-GO-Z	Other government sof nec	Other government software nec	7372
			7372
SOF-HL	Health services software	Software for the health services industry	7372
SOF-HL-C	Patient care mgmt sof	Software for specific types of diagnostic and therapeutic care	7372
SOF-HL-I	Integrated health ser sof	Integrated health services software	7372
SOF-HL-L	Laboratory management sof	Medical laboratory management software	7372
SOF-HL-M	Practice management sof	Health services practice management software	7372
SOF-HL-MA	Appointment scheduling sof	Medical appointment scheduling software	7372
SOF-HL-MB	Medical billing software	Medical billing software	7372
SOF-HL-MD	Dental practice mgmt sof	Software for dental practice management	7372
SOF-HL-MM	Medical practice mgmt softwa	Software for medical practice management	7372
SOF-HL-MZ	Other practice mgmt nec	Other practice management software nec	7372
SOF-HL-P	Pharmacy mgmt software	Software for pharmacy management	7372
SOF-HL-T	Institution mgmt software	Software for management of health/medical institutions	7372

(c) 1985 - 1994 Corporate Technology Information Services, Inc. All rights reserved.

SOF-HL-TH	Hospital admin/mgmt sof	Software for the management of hospitals	7372
SOF-HL-TM	Med personnel mgmt sof	Software for management of institutional personnel	7372
SOF-HL-TN	Nursing home mgmt sof	Software for the management of nursing homes	7372
SOF-HL-TP	Patient admin software	Software for the management of patient records	7372
SOF-HL-TZ	Other institutional sof	Other institutional management software nec	7372
SOF-HL-V	Veterinary software	Software for veterinary practices	7372
SOF-HL-Z	Other health services nec	Other health services software nec	7372
SOF-IN	Insurance software	Software for the insurance industry	7372
SOF-IN-A	Insurance management sof	Software to manage insurance agents and agencies	7372
SOF-IN-B	Actuarial software	Software to calculate life expectancies of insured persons	7372
SOF-IN-C	Claims processing software	Claims processing software	7372
SOF-IN-I	Integrated insurance sof	Integrated insurance software	7372
SOF-IN-P	Policy handling software	Policy handling software	7372
SOF-IN-Z	Other insurance software	Other insurance software nec	7372
SOF-LE	Legal software	Software for the legal service industry	7372
SOF-LE-A	Appointment scheduling sof	Legal appointment scheduling software	7372
SOF-LE-B	Legal time billing sof	Time billing software for legal practices	7372
SOF-LE-C	Litigation support sof	Software for legal litigation	7372
SOF-LE-D	Docket scheduling sof	Legal docket scheduling software	7372
SOF-LE-I	Integrated legal software	Software to manage legal practices	7372
SOF-LE-L	Legal database software	Software databases containing case abstracts or transcripts	7372
SOF-LE-M	Legislation tracking sof	Software to track legislation	7372
SOF-LE-T	Legal transcription sof	Software to enter court transcription	7372
SOF-LE-Z	Other legal software nec	Other legal software nec	7372
SOF-LI	Library software	Software to manage libraries	7372
SOF-MA	Manufacturing software	Manufacturing software systems	7372
SOF-MA-A	Capacity req planning sof	Software which determines utilization of manufacturers work centers for adjusting daily work schedules	7372
SOF-MA-B	Mfg automation protocol	Manufacturing automation protocol software	7372
SOF-MA-C	Robotic software	Robotic software	7372
SOF-MA-D	Operations planning sof	Operations planning software	7372
SOF-MA-E	Machine vision software	Machine vision software	7372
SOF-MA-F	Data acquisition/control	Factory data collection software	7372
SOF-MA-G	Eqp maintenance/mgmt sof	Software to track eqp & supplies maintenance histories	7372
SOF-MA-H	Group technology software	Group technology software	7372
SOF-MA-I	Integrated manufacturing sof	Integrated manufacturing software	7372
SOF-MA-L	Mfg resource planning sof	Manufacturing planning software	7372
SOF-MA-M	Material requirements sof	Material requirements software	7372
SOF-MA-MA	Material req planning sof	Material requirements planning software	7372
SOF-MA-MC	Bill of materials software	Bill of materials software	7372
SOF-MA-MZ	Other material req sof nec	Other material requirements software nec	7372
SOF-MA-N	Numerical control (NC) sof	Numerical control (NC) software	7372
SOF-MA-P	Process control software	Process control manufacturing systems software	7372
SOF-MA-Q	Quality control software	Software to control product quality	7372
SOF-MA-R	Production scheduling sof	Production scheduling software	7372
SOF-MA-S	Shop floor control sof	Shop floor control software	7372
SOF-MA-T	Test software	Test software	7372
SOF-MA-TC	Computer-aided testing sof	Computer-aided testing software	7372
SOF-MA-TD	Data reduction software	Data reduction software	7372
SOF-MA-TI	Test info integration sof	Test information integration software	7372
SOF-MA-TS	Test simulation software	Test simulation software	7372
SOF-MA-TT	Test software development	Test software development	7372
SOF-MA-TZ	Other test software nec	Other test software nec	7372
SOF-MA-U	Mfg simulation software	Manufacturing simulation software	7372
SOF-MA-W	Product service software	Software to manage/administer product services	7372
SOF-MA-Z	Other manufacturing nec	Other manufacturing software nec	7372
SOF-ME	Media communications sof	Media and communications software	7372
SOF-ME-A	CAR software	Computer-assisted retrieval (CAR) software	7372
SOF-ME-B	Book publishing software	Book publishing software	7372
SOF-ME-C	Circulation/subscription	Circulation and subscription management software	7372
SOF-ME-D	ADSTAR software	Automated data storage and retrieval (ADSTAR) software	7372
SOF-ME-E	Entertainment industry sof	Software for entertainment related industries	7372
SOF-ME-N	Newspaper/magazine sof	Newspaper and magazine publishing software	7372
SOF-ME-P	Electronic publishing sof	Software used to electronically create text for publication	7372
SOF-ME-T	Technical publication sof	Software for publishing technical or scientific publications	7372
SOF-ME-Z	Other media/communications	Other media/communications software nec	7372

SOF-NP	Non-profit org sof	Non-profit organization software	7372
SOF-NR	Natural resource mgmt software	Natural resource management software	7372
SOF-QA	Office automation sof	Office automation software	7372
SOF-QA-A	Appointment scheduling sof	Appointment scheduling software	7372
SOF-QA-D	Desktop management sof	Desktop management software	7372
SOF-QA-E	Data entry software	Data entry software	7372
SOF-QA-G	Graphics software	Graphics software	7372
SOF-QA-GA	Animation software	Animation graphics software	7372
SOF-QA-GB	Business graphics software	Business graphics software	7372
SOF-QA-GD	3D software	Three dimensional representation software	7372
SOF-QA-GI	Image processing software	Image processing software	7372
SOF-QA-GM	Biomedical graphics sof	Biomedical graphics software	7372
SOF-QA-GS	Slide making software	Slide making software	7372
SOF-QA-GV	Videotex software	Videotex software	7372
SOF-QA-GZ	Other graphics software	Other graphics software nec	7372
SOF-QA-H	Human resource mgmt sof	Software to manage human resources	7372
SOF-QA-I	Idea management software	Idea management software	7372
SOF-QA-L	Mailing list management	Mailing list management software	7372
SOF-QA-M	Electronic mail software	Electronic mail software	7372
SOF-QA-MB	Elttrnc bulletin board sof	Electronic bulletin board software	7372
SOF-QA-MC	Comp teleconferencing sof	Computer teleconferencing software	7372
SOF-QA-ME	Elttrnc message systems sof	Electronic message systems software	7372
SOF-QA-MG	Groupware	Software that helps groups work more productively	7372
SOF-QA-MV	Voice message systems sof	Voice message systems software	7372
SOF-QA-MZ	Other electronic mail sof	Other electronic mail software nec	7372
SOF-QA-P	Desktop publishing sof	Desktop publishing software	7372
SOF-QA-S	Spreadsheet software	Standalone spreadsheet packages and integrated software products with a spreadsheet core	7372
SOF-QA-T	Time management software	Time management software	7372
SOF-QA-W	Word processor/text editor	Word processor/text editor software	7372
SOF-QA-Z	Other office automation	Other office automation software nec	7372
SOF-PD	Program development sof	Software development systems	7372
SOF-PD-A	Cross assembler software	Software to translate assembler source code from for one computer to machine language code for another computer	7372
SOF-PD-D	Portable development sof	Programming language that is implemented on multiple computer systems	7372
SOF-PD-E	Machine code linkers	Software which links machine code programs	7372
SOF-PD-F	Code algorithms	Procedures to arrive at solutions for problems	7372
SOF-PD-G	Lex analyzer generators	Lexical analyzer generators	7372
SOF-PD-L	Progrmg lang enhancements	Enhancements, supersets, and extensions to programming languages	7372
SOF-PD-N	Language development sof	Language development software	7372
SOF-PD-NA	Language assemblers	Software to translate from assembler or macro language to machine code	7372
SOF-PD-NC	Language compilers	Software to translate from high level languages to machine languages	7372
SOF-PD-NCA	Ada	Compilers/interpreters for the high level language Ada	7372
SOF-PD-NCB	BASIC	Compilers/interpreters for the high level language BASIC	7372
SOF-PD-NCC	C language	Compilers/interpreters for the high level language C	7372
SOF-PD-NCD	Pascal	Compilers/interpreters for the high level language Pascal	7372
SOF-PD-NCF	FORTH	Compilers/interpreters for the high level language FORTH	7372
SOF-PD-NCH	Modula-2	Compilers/interpreters for the high level language Modula-2	7372
SOF-PD-NCO	COBOL	Compilers/interpreters for the high level language COBOL	7372
SOF-PD-NCP	APL	Compilers/interpreters for the high level language APL	7372
SOF-PD-NCT	FORTRAN	Compilers/interpreters for the high level language FORTRAN	7372
SOF-PD-NCZ	Other compilers	Other compilers/interpreters nec	7372
SOF-PD-NZ	Other language devel sof	Other language development software nec	7372
SOF-PD-O	Code loaders	Software to alter machine code relocatable addresses, and store the code in memory	7372
SOF-PD-P	Parser generators	Parser generators	7372
SOF-PD-R	Preprocessors	Software to produce input to compilers	7372
SOF-PD-T	Program translators	Software to translate one high level language to another	7372
SOF-PD-X	Cross compilers	Compiler running on one computer systems that generates object code to run on a different system	7372
SOF-PD-Z	Other software devel	Other software development systems nec	7372
SOF-PM	Project management sof	Project management software	7372

SOF-PU	Public utilities software	Software for public utilities	7372
SOF-RE	Real estate software	Software to manage properties	7372
SOF-SM	Sales/marketing software	Sales and marketing software	7372
SOF-SM-A	Sales analysis software	Sales analysis/reporting software	7372
SOF-SM-D	Direct marketing software	Direct marketing software	7372
SOF-SM-I	Integrated marketing sof	Integrated marketing software	7372
SOF-SM-R	Survey analysis software	Market survey and research analysis software	7372
SOF-SM-T	Telemarketing mgmt sof	Telemarketing management software	7372
SOF-SM-Z	Other sales/marketing sof	Other sales and marketing software nec	7372
SOF-SR	Service industry software	Software for services and service industries	7372
SOF-SV	Software services	Software-related services	7370
SOF-SV-A	Artificial intel services	Services for or relating to artificial intelligence	7370
SOF-SV-AC	AI consulting services	Artificial intelligence consulting services	7379
SOF-SV-AP	AI software programming	Custom AI software programming services	7371
SOF-SV-AR	AI R&D services	Artificial intelligence R&D	7371
SOF-SV-AZ	Other AI services nec	Other artificial intelligence (AI) services nec	7379
SOF-SV-C	Software consulting ser	Software consulting services	7379
SOF-SV-P	Custom sof programming ser	Custom computer programming services	7371
SOF-SV-PA	Applications software	Custom applications software programming services	7371
SOF-SV-PS	Systems software	Custom systems software programming services	7371
SOF-SV-PZ	Other custom programming	Other custom software programming services nec	7371
SOF-SV-Z	Other software services	Other software-related services nec	7379
SOF-TR	Transportation software	Software for transportation industries	7372
SOF-TR-A	Aerospace software	Software for aerospace and aviation	7372
SOF-TR-R	Transportation reserv sof	Software for reservations in transportation industries	7372
SOF-TR-V	Vehicle sales/service softwar	Software for vehicle sales/service industries	7372
SOF-TR-Z	Other transportation sof	Other transportation software nec	7372
SOF-TS	Technical/scientific sof	Technical/scientific software	7372
SOF-TS-E	Engineering/technical sof	Software for the engineering and technical industries	7372
SOF-TS-EA	Chem engineering/tech sof	Chemical engineering software	7372
SOF-TS-EC	Civil/struct/arch eng sof	Civil/structural/architectural software	7372
SOF-TS-EE	Electrical/eltrnc eng sof	Electrical/electronic engineering software	7372
SOF-TS-EF	Mechanical engineering sof	Mechanical software	7372
SOF-TS-EG	General purpose eng sof	General purpose engineering software	7372
SOF-TS-EM	Manufacturing eng/tech sof	Manufacturing engineering software	7372
SOF-TS-ER	Geographic information sys	Geographic information systems software	7372
SOF-TS-EU	Structural analysis sof	Software used for structural analysis	7372
SOF-TS-EV	Software eng/technical sof	Software engineering software	7372
SOF-TS-EZ	Other engineering/tech nec	Other engineering and technical software nec	7372
SOF-TS-T	Scientific software	Software for scientific functions	7372
SOF-TS-TA	Physical sciences software	Software for laboratory and experiment processes in the physical sciences	7372
SOF-TS-TH	Math and stats software	Software for mathematical or statistical functions	7372
SOF-TS-TZ	Other scientific software	Other scientific software nec	7372
SOF-UT	Utility systems software	Utility systems software	7372
SOF-UT-5	Data conversion software	Software to convert data from one format to another	7372
SOF-UT-A	Program generators	Software that automates the development of programs	7372
SOF-UT-B	Screen formatting software	Software to design CRT screen layouts and define data entry fields	7372
SOF-UT-C	Debugging & testing sof	Debugging and testing software	7372
SOF-UT-E	Operating sys enhancements	Extensions to standard operating systems	7372
SOF-UT-F	Operating systems	Operating systems	7372
SOF-UT-G	Windowing systems	Software to allow access to multiple, unrelated applications simultaneously	7372
SOF-UT-H	Peripheral device drivers	Software to control the operation of peripheral devices	7372
SOF-UT-I	Print utilities/spoolers	Software to control printing operations, provide special print utilities and handle spooling of data to hardcopy devices	7372
SOF-UT-J	Disk/tape/file utilities	Software to facilitate the use of off-line storage devices	7372
SOF-UT-K	Font utilities	Font utilities, letter	7372
SOF-UT-L	Librarian utilities	Software to log and track the location of data files	7372

(c) 1985 - 1994 Corporate Technology Information Services, Inc. All rights reserved.

SOF-UT-M	Graphics utilities	Graphics utilities	7372
SOF-UT-N	System design/methodology	Software to assist in the control of large system development projects	7372
SOF-UT-O	Data center mgmt software	Software to control program execution and monitor system usage	7372
SOF-UT-P	Programming utilities	Packages used for the automation of routine programming tasks	7372
SOF-UT-Q	Application devel tools	Tools and aids which provide control over all phases of software development	7372
SOF-UT-R	Disaster recovery software	Software to reestablish a system's data after an unscheduled interruption	7372
SOF-UT-S	Sort/merge sof utilities	Software to re-order one or more data files	7372
SOF-UT-T	Documentation generators	Software to create documentation for programs	7372
SOF-UT-U	Menu/screen facilities	Software to generate CRT menus and input screens	7372
SOF-UT-V	Word processing utilities	Extensions to word processing packages	7372
SOF-UT-W	Report generators	Software to facilitate the process of formatting printed reports	7372
SOF-UT-X	Security/auditing software	Software control access to the system and sensitive files, and to verify a file's contents	7372
SOF-UT-Y	Performance measuring sof	Software to provide information to fine tune and control a system's performance	7372
SOF-UT-Z	Other utility software nec	Other utility software nec	7372
SOF-WD	Warehousing/distribution	Warehousing and distribution software	7372
SOF-WD-D	Distribution mgmt software	Distribution management software	7372
SOF-WD-F	Freight/warehouse software	Freight, warehousing and trucking software	7372
SOF-WD-FD	Freight dispatch software	Software to dispatch vehicles for transportation or public service	7372
SOF-WD-FF	Freight handling software	Freight handling software	7372
SOF-WD-FR	Route scheduling software	Software to schedule routes for vehicles	7372
SOF-WD-FT	Truck/fleet mgmt sof	Truck/fleet management software	7372
SOF-WD-FW	Warehouse/storage mgmt sof	Warehouse/storage management software	7372
SOF-WD-FZ	Other freight/warehousing	Other freight/warehousing/trucking software nec	7372
SOF-WD-I	Inventory management sof	Inventory management software	7372
SOF-WD-O	Order entry/processing sof	Order entry/processing software	7372
SOF-WD-Z	Other distribution sof nec	Other warehousing/distribution software nec	7372
SOF-ZA	Applications software nec	Other applications software nec	7372
SOF-ZA-C	Trade software	Software for trade organizations	7372
SOF-ZA-CI	Import/export trade sof	Import/export trade software	7372
SOF-ZA-CR	Retail trade software	Retail trade software	7372
SOF-ZA-CW	Wholesale trade software	Wholesale trade software	7372
SOF-ZA-CZ	Other trade software nec	Other trade software nec	7372
SOF-ZA-F	Foreign lang translation	Software to translate from one non computer language to another	7372
SOF-ZA-G	General simulation sof	Software which simulate environments and conditions	7372
SOF-ZA-H	Home use software	Software designed for the consumer marketplace	7372
SOF-ZA-X	Special needs software	Software to meet the needs of handicapped individuals	7372
SOF-ZA-Z	Other applications nec	Other applications software nec	7372
SOF-ZZ	Other software	Other software nec	7379

US Geological Survey



BIOSIS

Category Indexing Schemes: A BIOSIS Perspective

CENDI Indexing Workshop
September 21, 1994

Presented by Colleen Finley, BIOSIS

CATEGORY INDEXING SCHEMES: A BIOSIS PERSPECTIVE

OVERVIEW

The BIOSIS Previews Database covers the life sciences from the traditional fields of biology to interdisciplinary fields and a wide range of medically-oriented subjects. BIOSIS covers material from over 7000 journals, books and meetings through our major print publications, Biological Abstracts (BA) and Biological Abstracts/Reports, Reviews, Meetings (BA/RRM), and our online database, BIOSIS Previews. Currently BIOSIS Previews contains over 9.3 million records and will add 540,000 items in 1994.

Because of the diverse range of information that falls under the category of "life sciences information", BIOSIS found a need to partition this material into smaller, more manageable subsets or "Categories". To accomplish this, BIOSIS developed a list of core concepts covering a wide range of subject areas called Concept Codes. Since their introduction in 1969, Concept Codes have become an integral part of our indexing at BIOSIS.

The focus of my talk today is on Category Indexing Schemes. In particular, I will look at BIOSIS's experience with this type of indexing through our Concept Codes. I will discuss several aspects of Concept Code Indexing and use including:

- The role of Concept Codes in BIOSIS's general indexing scheme
- BIOSIS Indexing Policies for Concept Codes including how our staff assign these codes, our training guidelines, and the tools used to help indexers assign these codes consistently
- How Concept Codes appear in BIOSIS products and how they are used by our customers and staff
- Our experiences with Concept Code indexing, both positive and negative
- A look at what the future holds for 'category-based' indexing at BIOSIS
- Some points to consider when implementing a 'category-based' indexing scheme

BIOSIS SUBJECT COVERAGE

► Traditional Fields of Biology

- Botany
- Zoology
- Genetics
- Entomology
- Ecology
- Microbiology
- Anatomy
- Embryology

► Interdisciplinary Fields

- Clinical Microbiology
- Biophysics
- Pharmacology
- Biochemistry
- Biotechnology
- Paleobiology

► Medical Subjects

- Clinical Research
- Surgery
- Preclinical Research
- Applied Research
- Veterinary Medicine
- Forensic Medicine

BIOSIS MAJOR PRODUCTS

► Print

- Biological Abstracts (BA)
- Biological Abstracts/RRM (Reports, Reviews, Meetings) (BA/RRM)
- Zoological Record

► Online Products

- BIOSIS Previews
- BioBusiness
- Zoological Record Online

► CD-ROM Products

- Biological Abstracts on Compact Disk
- Biological Abstracts/RRM on Compact Disk
- Zoological Record on Compact Disk
- Foods Intelligence on Compact Disk

BIOSIS INDEXING TERMS

- ▶ **Natural Language (Free Text)**
 - Majority of BIOSIS keywords
 - Includes organism names, substances, procedures, methods...

- ▶ **Controlled Keywords**
 - Drug Affiliations
 - Geographic Terms
 - Taxonomic Affiliations
 - Bibliographic Descriptors
 - Subject Specific Areas
(ex: Molecular Sequence Data)

- ▶ **Biosystematic Codes (BC)**
 - Five-digit codes used to index organisms
 - Over 900 codes representing taxonomic categories
 - Searchable by code or name on most systems
 - Allows searching by Super Taxonomic Groups

- ▶ **Concept Codes (CC)**
 - Unique five digit codes representing broad subject areas in the life sciences
 - Over 550 Available Concept Codes
 - Searchable by Code or Name on most systems

CONCEPT CODE NUMERIC DIRECTORY

SAMPLE ENTRY

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
GENETICS AND CYTOGENETICS	CC03502 CC03504 CC03506 CC03508 CC03509 CC03510	General Plant Animal Human Population Genetics Sex Differences	Introduced in 1972

- ▶ **Broad Concept Heading**
 - Broad Subject Category used to group related codes and headings
- ▶ **Concept Code**
 - Five digit code assigned by BIOSIS
 - Preceded by CC to distinguish from Biosystematic Codes (BC)
- ▶ **Concept Heading**
 - Specific Concept title
- ▶ **Notes**
 - Provides historical information about Concept Code

CONCEPT CODE ASSIGNMENT (WEIGHTING)

► Primary Assignment

- Represents the major focus of the paper
- Determines the Major Subject Heading and Subheading where item will appear in print product
- Only one primary code per source document

► Secondary Assignment

- Identifies additional subject areas of major importance in the source document
- For electronic products, primary and secondary codes become major level codes

► Tertiary Assignment

- Assigned to concepts mentioned in the source document but not the major emphasis of the paper
- For electronic products, becomes minor level code

INDEXING TOOLS AND REFERENCES

► Online Indexing Tools

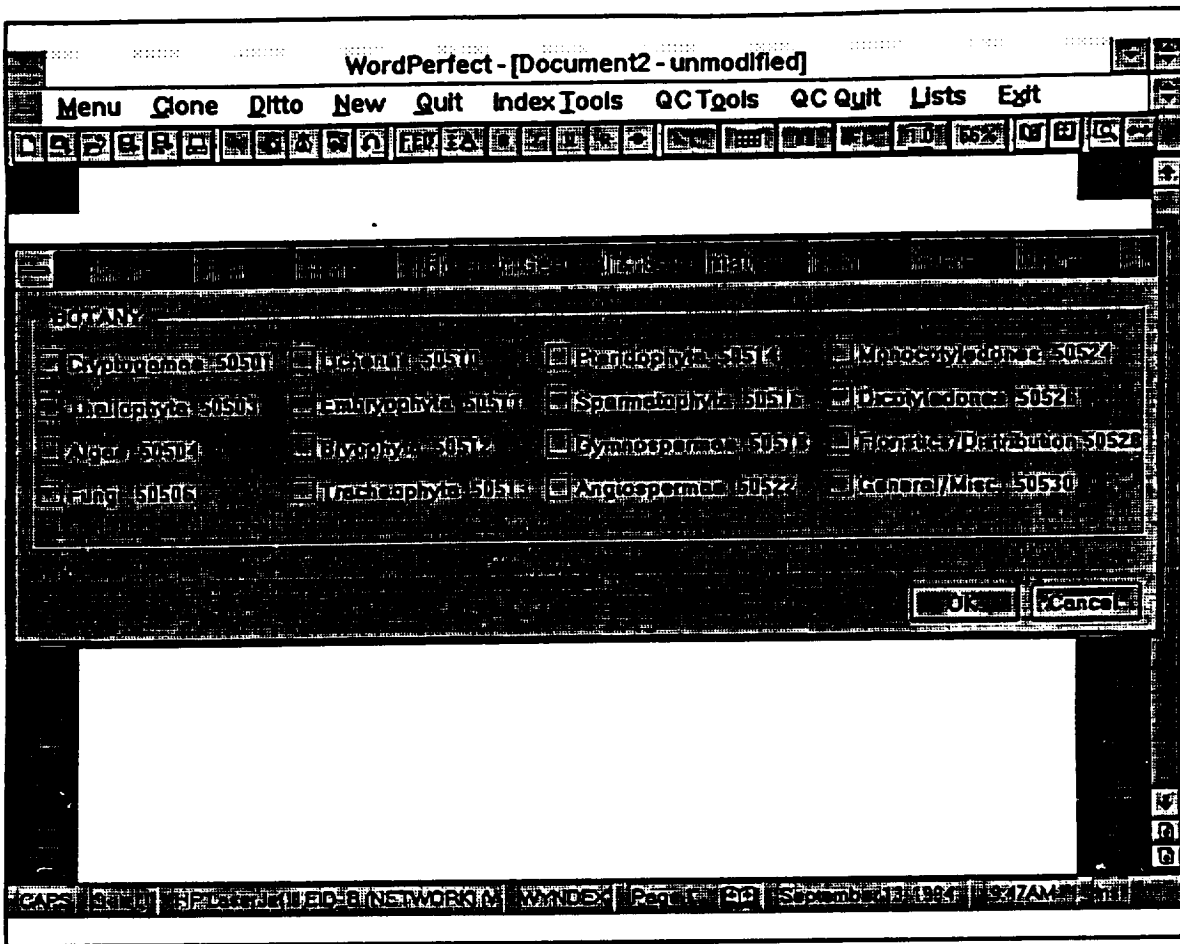
- **Concept Code Authority File**
 - Lists codes and code name
- **Subject Specific lists of Concept Codes**
- **Subject-Specific Style Sheets**
 - Includes pick lists of Concept Codes
- **Interactive Validation Checks**
 - Primary Assignment
 - Valid Concept Code

► Print References and Other Resources

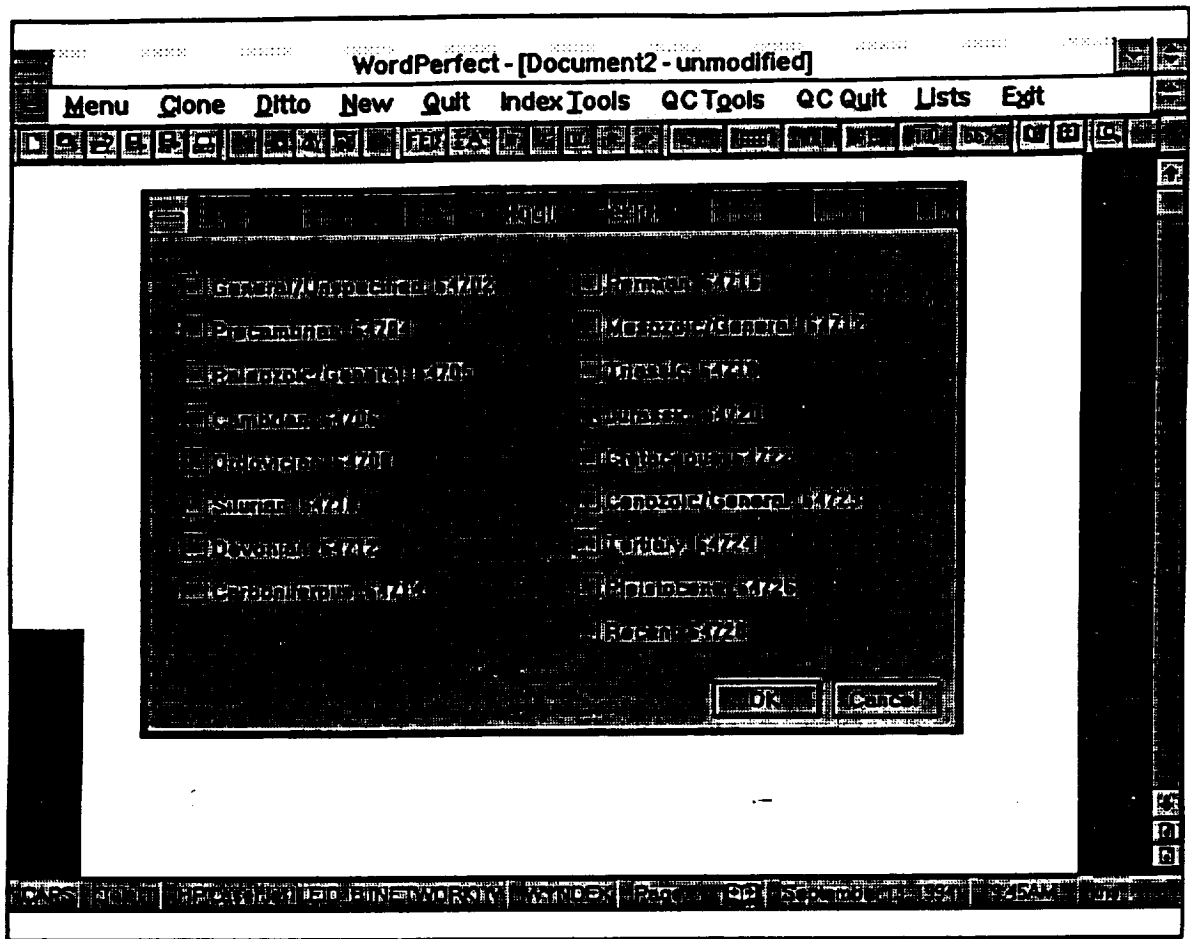
- **Supplemental Information for Indexing**
- **Section Definitions for Cross Indexing**
- **In-house database containing one year of BIOSIS Previews**
- **Scope Notes from the BIOSIS Previews Search Guide**

Move Codes	Move Abbreviation	Move Drug Affiliation	Move Geography	Ext
CONCEPT CODES				
C00502	GENERAL BIOLOGY - PHILOSOPHY			
C00504	GENERAL BIOLOGY - TAXONOMY, NOMENCLATURE AND TERMINOLOGY			
C00506	GENERAL BIOLOGY - EXPLORATIONS, EXPEDITIONS, ETC.			
C00508	GENERAL BIOLOGY - INSTITUTIONS, ADMINISTRATION AND LEGISLATION			
C00510	GENERAL BIOLOGY - MUSEUMS, BOTANICAL/ZOOLOGICAL GARDENS, AQUARIA			
C00512	GENERAL BIOLOGY - CONSERVATION, RESOURCE MANAGEMENT			
C00514	GENERAL BIOLOGY - TEXTBOOKS, GENERAL; EDUCATION; AUDIO-VISUAL AIDS			
C00516	GENERAL BIOLOGY - NATURE STUDY AND MATERIALS			
C00518	GENERAL BIOLOGY - Nontechnical Literature (Minor Only)			
C00520	GENERAL BIOLOGY - SYMPOSIA, TRANSACTIONS, PROCEEDINGS			
C00522	GENERAL BIOLOGY - HISTORY AND ARCHAEOLOGY			
C00524	GENERAL BIOLOGY - BIOGRAPHIES			
C00526	GENERAL BIOLOGY - BIBLIOGRAPHY			
C00530	GENERAL BIOLOGY - INFORMATION, DOCUMENTATION, RETRIEVAL AND COMPUTER APPLICATIONS			
C00531	GENERAL BIOLOGY - FORENSIC SCIENCE			
C00532	GENERAL BIOLOGY - MISCELLANEOUS			
C01004	METHODS - LABORATORY METHODS			
C01006	METHODS - LABORATORY APPARATUS			
C01008	METHODS - FIELD METHODS			
C01010	METHODS - FIELD APPARATUS			
C01012	METHODS - PHOTOGRAPHY			
C01014	METHODS - APPARATUS - MISCELLANEOUS			
C01052	MICROSCOPY - GENERAL AND SPECIAL TECHNIQUES			
C01054	MICROSCOPY - CYTOLOGY AND CYTOCHEMISTRY			
C01056	MICROSCOPY - HISTOLOGY AND HISTOCHEMISTRY			
C01058	MICROSCOPY - ELECTRON MICROSCOPY			
C01500	EVOLUTION			
C02502	CYTOLOGY & CYTOCHEMISTRY - GENERAL			
C02504	CYTOLOGY & CYTOCHEMISTRY - PLANT			
C02506	CYTOLOGY & CYTOCHEMISTRY - ANIMAL			
C02508	CYTOLOGY & CYTOCHEMISTRY - HUMAN			
C03502	GENETICS & CYTOGENETICS - GENERAL			
C03504	GENETICS & CYTOGENETICS - PLANT			
C03506	GENETICS & CYTOGENETICS - ANIMAL			

ON-LINE CONCEPT CODES AUTHORITY FILE



PICK LIST OF PLANT SYSTEMATICS CONCEPT CODES



CONCEPT CODE PICK LIST OF GEOLOGICAL PERIODS

ORIGINAL PAGE IS
OF POOR QUALITY

Pharmacology:

Drug:

Admission:

Blood
 HEMATOLOGIC-DRUG:
 ANTICOAGULANT-DRUG:
 ANTISICKLING-DRUG:
 REPLENISHING AGENT-DRUG:
 THROMBOLYTIC-DRUG:

Type of Study:
☐ Animal Study
☐ Practical Test
☐ Clinical Trial

Organism:
☐ Human
☐ Rat
☐ Mouse
☐ Guinea Pig
☐ Rabbit
☐ Dog
☐ Cat
☐ Monkey
☐ Rhesus Monkey

Other:

Concept Code:
☐ General - 22002
☒ Pharmacology - 22003
☒ Clinical - 22005
☐ Blood - 22001
☐ Cardiovascular - 22010
☐ Connective Tissue - 22012
☐ Digestive System - 22014
☐ Endocrine System - 22011
☒ Immunology/Allergy - 22013
☐ Integumentary System - 22020
☐ Muscles - 22024
☐ Neuropharmacology - 22026
☐ Psychopharmacology - 22028
☐ Reproductive System - 22021
☐ Respiratory System - 22030
☐ Sensory Organ - 22031
☐ Urinary System - 22032

☐ Route of Administration
☐ Suggested Keyword

OK Cancel

PCAS: Pharmacology Style Sheet - WORKING - WYNDEX - Page 19 - September 1994 - 1:52AM

PHARMACOLOGY STYLE SHEET WITH PICK LIST OF PHARMACOLOGY CONCEPT CODE SUBSET

ORIGINAL PAGE 18
 OF POOR QUALITY

SUPPLEMENTAL INFORMATION FOR INDEXING

ALL CROSSES MAY BE USED AT ANY APPROPRIATE INDEXING LEVEL

127

SUBJECT TERM	AUTOMATIC CROSS(ES)	COMMON CROSS(ES)	TAXONOMY
Hereditary fructose intolerance	0350(8) 13020 10808 25503 13004	10064 13012 10068 13229	
Hereditary hemolytic anemia	0350(8) 15006 13020	0250- 15004	
Hereditary hemorrhagic telangiectasia	03508 14508		
Hereditary methemoglobinemias see Methemoglobinemias, hereditary or congenital			
Hereditary nonhemolytic bilirubinemia	0350(8) 13020 13013 14006	10065 15002	
Hereditary nonspherocytic hemolytic anemia	0350(8) 15006 10808 25503 13020	02508 13012 10064 15004	
Hereditary spherocytosis	0350(8) 25503 15006	02508 15004 14006 15008	
Hereditary unconjugated hyperbilirubinemia see Hyperbilirubinemia, (hereditary) unconjugated			
Heredopathia atactica polyneuritiformis (Refsum's disease)	0350(8) 13020 10808 20506 13006	10064 13012 10066	
Hering-Breuer reflex	16004 20504		
Hermaphroditism	1650-	03510 25552 1700(4,6)	

SECTION DEFINITIONS FOR CROSS INDEXING

GENETICS OF BACTERIA (cont.)/GENETICS AND CYTOGENETICS

(31500) GENETICS OF BACTERIA AND VIRUSES (cont.)

Excludes: DNA as a chemical constituent [see Morphology and Cytology of Bacteria; Physiology and Biochemistry of Bacteria].

Physiologic repression; non-genetically controlled feedback inhibition of biosynthesis of enzymes or other chemical constituents [see Physiology and Biochemistry of Bacteria].

(03500) GENETICS AND CYTOGENETICS

[Organisms studied include Plants, Invertebrates and Chordates]

Includes: Biochemical and physiological studies of genetic interest (transmission of genetic information; DNA replication; genetic-oriented cellular studies of DNA and RNA, i.e., messenger (information) and transfer (soluble) RNA; enzyme differences resulting from gene variations, etc.).

Chromosome counts, mapping and crossing over, aberrations (e.g., deletions, duplications, inversions, translocations, aneuploids, polyploids, etc.) and chromosome basis for pathology.

Mechanisms of genetic change including mutagenic agents (drugs, chemical, mechanical, radiation, etc.), mutation studies (germinal and somatic) and hybrids and the effects of hybridization.

Genotypic and phenotypic expression and effects of polyploidy.

Extra chromosomal inheritance (plastids, plasmagones).

Genetics of blood groups, inheritance of blood-group factors or secretory traits and immunogenetic studies.

Heritable and familial anomalies and pathological occurrences (e.g., metabolic disturbances, hormonal disturbances, congenital anomalies, etc.) and their developmental basis.

Excludes: Bacterial and viral genetics except when part of a mixed study involving animals and/or plants [see Genetics of Bacteria and Viruses].

Structural studies of DNA and RNA [see "Molecular Properties and Macromolecules" under Biophysics or "Plant Cytology" if plant study].

(03506) ANIMAL

Includes: All genetic and cytogenetic studies relating to invertebrates and chordates (including mixed animal and human studies).

Excludes: Human genetics [see "Human" under Genetics and Cytogenetics].

USER SEARCH REFERENCES

► ***How To Search Series***

- Biological Abstracts on Compact Disk and Biological Abstracts/RRM on Compact Disk
- Biological Abstracts and Biological Abstracts/RRM
- BIOSIS Previews

► ***Discovering BIOSIS Previews: An Interactive Disk and Guide***

► **BIOSIS Previews Search Guide**

- Concept Code Directory
 - Numerical List of all Concept Codes
- Master Index
 - Lists 19,500 terms and phrases with recommendations for searching on BIOSIS Previews
- Concept Code Scope Notes
 - Defines Concept Codes and provides examples of scope and coverage

MASTER INDEX

SAMPLE ENTRY

GENETIC

EX: *GENETIC*\$ (54860)
see also GENE; GENIC; GENOTYPE;
HEREDITARY; HERITABILITY;
INHERITANCE

CC: GENETIC

Animal Genetics and Cytogenetics

CC03506 (121550)

Genetic Sex Differences

CC03510 (34240)

Genetics and Cytogenetics

CC03502 (16930)

Genetics of Bacteria and Viruses

CC31500 (74700)

Human Genetics and Cytogenetics

CC03508 (128620)

Plant Genetics and Cytogenetics

CC03504 (60900)

Population Genetics

1972- : ***CC03509*** (13830)

* Numbers in Parentheses are five year frequency counts

SCOPE NOTES

SAMPLE ENTRY

GENETICS AND CYTOGENETICS CC03502-CC03510

The codes in this group retrieve studies on the origin and transmission of inherited characteristics, genetic variations and sex determination in plants, animals and humans.

Examples Studies on * genetic engineering * chromosome mapping * mutagens

Strategy Recommendations

- * For radiation mutagenesis studies, use relevant *Genetics and Cytogenetics* codes with appropriate keywords and the *Radiation Effects and Protective Measures* code CC06506.

*

*

*

CC03502 GENERAL

Frequencies Major (15570) Minor (1360)

Applications This code retrieves general studies that cannot be accessed by one of the other *Genetics and Cytogenetics* codes. It also accesses information on the apparatus and methods used in genetic studies.

Examples Studies on * evolution of the genetic code * general genetic modeling * mutation rates

Strategy Recommendations

- * For mathematical models useful in genetic studies, use this code with appropriate keywords, the *Mathematical Biology and Statistical Methods* code CC04500 and the *Biocybernetics* code CC10515.

CONCEPT CODE SCOPE NOTES

A total of 571 Concept Codes are used to index broad subject areas in the BIOSIS Previews database. The Concept Code Scope Notes section presents these Concept Codes in numerical order, defines them, and provides examples of their subject scope and coverage. Use the Concept Code Scope Notes to learn how Concept Codes can be used to make your searching more effective and economical.

①

TOXICOLOGY CC22501-CC22508

BROAD CONCEPT HEADING
(includes applicable Concept
Code or Concept Code range)

The codes in this group retrieve studies on toxic substances including their actions and detection in plants, animals and humans. They access information on antidotes and treatments used to alleviate or prevent toxic reactions.

Examples Studies on • animal venoms • poisonous gases • drug toxicity • endotoxins • heavy metal intoxication • phytotoxins • antidotes • pesticide poisoning • fluoride toxicity

Strategy Recommendations

- For studies of the toxic effects of a substance on organ system(s), use relevant *Toxicology* codes with appropriate keywords and specific organ system *Pathology* codes.

BROAD CONCEPT DESCRIPTION
(occurs only for Broad Concept
Headings encompassing a
range of codes)

②

CC22506 ENVIRONMENTAL AND INDUSTRIAL TOXICOLOGY

③

Frequencies Major (56280) Minor (5370)

④

Applications This code retrieves studies on exposure to and effects of toxic substances in living and working environments.

⑤

Examples Studies on • industrial toxicants • toxic agents in drinking water • occupational poisons • asbestosis • toxic fumes • cadmium exposure toxicity

⑥

Strategy Recommendations

- For toxicological studies in plants, use this code with the *Phytopathology-Nonparasitic Diseases* code CC54512 and the relevant *Plant Physiology, Biochemistry and Biophysics* codes.
- For studies of occupational exposure to hazardous substances, use this code with the *Occupational Health* code CC37013.
- For environmental radioactive chemical toxicity studies, use this code with the *Radiation Effects and Protective Measures* code CC06506 and the *Public Health: Environmental Health-Radiation Health* code CC37017.
- For toxicology studies of freshwater organisms, use this code with appropriate keywords and the *Limnology* code CC07514.
- For studies of the toxic agents in air, water or soil, use this code with appropriate keywords and the *Air, Water and Soil Pollution* code CC37015.

INDIVIDUAL SCOPE NOTE

⑦

Coverage Note In 1972, the scope of this code was expanded to include environmental toxicology.

1. BROAD CONCEPT HEADING AND CODE RANGE

Concept Codes are grouped under Broad Concept Headings or subject areas. These headings correspond to those used in the Concept Code Directory. The code or range of codes comprising the subject area is listed for each heading. When a range of codes is indicated, a general description of the subject scope and coverage of the section, including examples and strategy recommendations, is given.

3. FREQUENCIES

These frequencies are the estimated number of documents indexed at the Major and Minor levels for each Concept Code for the five-year period 1988 through 1992. Major Concept Codes are used to indicate the ideas of primary interest discussed in a paper. Topics of lesser importance or emphasis are indexed at the Minor level. A more detailed discussion of indexing levels is given in the Introduction to the Concept Code Directory.

2. CONCEPT CODE AND CONCEPT HEADING

Each Concept Code Scope Note begins with the Concept Code, identified by "CC" preceding a five-digit number, followed by its 'Concept Heading' or title.

4. APPLICATIONS

A broad statement of the subject area and organisms covered by each Concept Code is given.

EXAMPLES

Examples of specific topics included in the scope are listed here to help clarify the subject areas covered by the Concept Code. These examples have been provided as a guideline to help in selecting appropriate Concept Codes for your search strategies. These samples highlight many, but not every, topic included in the scope of the Concept Code.

STRATEGY RECOMMENDATIONS

Subject-specific recommendations are given to show how the Concept Codes can be used with each other and with other BIOSIS indexing elements (i.e., Biosystematic Codes

or Added Keywords) to construct effective and relevant search strategies. Additional information concerning these indexing elements can be found in the Master Index and the Biosystematic Code Directory.

7. COVERAGE NOTE

Included only for certain Concept Codes, this note provides details about changes in and limits of the scope of the Code. While BIOSIS Previews may contain some citations on topics not generally included in the scope of the Code, searchers may wish to consult other databases for comprehensive retrieval. For Concept Codes that have been introduced or undergone a change in scope since 1969, the year of the change is listed here.

- For cell chemistry studies in humans, use this code with appropriate keywords and relevant *Biochemical Studies* codes.
- For drug studies in human cells, use this code with appropriate keywords and relevant *Pharmacology* codes.

GENETICS AND CYTOGENETICS CC03502-CC03510

The codes in this group retrieve studies on the origin and transmission of inherited characteristics, genetic variations and sex determination in plants, animals and humans.

Examples Studies on • genetic engineering • chromosome mapping • mutagens

Strategy Recommendations

- For radiation mutagenesis studies, use relevant *Genetics and Cytogenetics* codes with appropriate keywords and the *Radiation Effects and Protective Measures* code CC06506.
- For studies on the genetic aspects of evolution, use relevant *Genetics and Cytogenetics* codes with appropriate keywords and the *Evolution* code CC01500.
- For genetic studies of the developing embryo, use relevant *Genetics and Cytogenetics* codes with appropriate keywords and *Developmental Biology; Embryology* codes.
- For bacterial or viral genetics studies, use the *Genetics of Bacteria and Viruses* code CC31500.
- For studies on DNA replication, use the *Replication, Transcription and Translation* code CC10300.

CC03502 GENERAL

Frequencies Major (15570) Minor (1360)

Applications This code retrieves general studies that cannot be accessed by one of the other *Genetics and Cytogenetics* codes. It also accesses information on the apparatus and methods used in genetic studies.

Examples Studies on • evolution of the genetic code • general genetic modeling • mutation rates

Strategy Recommendations

- For mathematical models useful in genetic studies, use this code with appropriate keywords, the *Mathematical Biology and Statistical Methods* code CC04500 and the *Biocybernetics* code CC10515.

CC03504 PLANT

Frequencies Major (57000) Minor (3900)

Applications This code retrieves genetic and cytogenetic studies of plants.

Examples Studies on • hybridization • chromosome structure • karyological research • genetic resistance to disease

Strategy Recommendations

- For plant hybridization and agricultural yield studies, use this code with appropriate keywords and relevant *Agronomy or Horticulture* codes.
- For genetic studies of resistance to plant disease, use this code with appropriate keywords and the *Phytopathology-Parasitism and Resistance* code CC54514.
- For radiation-induced genetic effects in plants, use this code with appropriate keywords and the *Plant Physiology-Light and Radiation Effects* code CC51516.
- For genetic studies of plant taxonomy, use this code with appropriate keywords and the relevant *Botany, General and Systematic* codes.

CC03506 ANIMAL

Frequencies Major (113460) Minor (8090)

Applications This code retrieves nonhuman animal genetic and cytogenetic studies.

Examples Studies on • mutations in fish • mutagenicity screening in mice • genetic control of disease susceptibility • DNA repair in mammalian cells

Strategy Recommendations

- For genetics of specific animal organ systems, use this code with appropriate keywords and relevant organ system codes.
- For environmental chemicals that are mutagenic to animals, use this code with appropriate keywords and the *Environmental and Industrial Toxicology* code CC22506.
- For radiation-induced chromosomal aberrations in animals, use this code with appropriate keywords and the *Radiation Effects and Protective Measures* code CC06506.
- For genetic characteristics of domestic and economically important animals including poultry, use this code with appropriate keywords and the *Animal Production-Breeds and Breeding* code CC26506 or the *Poultry Production-Breeds and Breeding* code CC27006.

CC03508 HUMAN

Frequencies Major (118810) Minor (9810)

Applications This code retrieves all genetic and cytogenetic studies of humans.

Examples Studies on • chemical mutagenesis in humans • sister chromatid exchanges in human cells • genetic counseling • genetic markers • genetic analysis of human diseases

Strategy Recommendations

- For genetic studies of human evolution, use this code with appropriate keywords and the *Evolution* code CC01500.
- For genetic studies of metabolic disorders, use this code with appropriate keywords and the *Metabolic Disorders* code CC13020.
- For mutagen-induced cancer in humans, use this code with appropriate keywords and the *Carcinogens and Carcinogenesis* code CC24007.
- For genetic studies used to describe racial variations, use this code with appropriate keywords and the *Physical Anthropology; Ethnobiology* code CC05000.
- For inherited susceptibilities to cancer, use this code with appropriate keywords and relevant *Neoplasms and Neoplastic Agents* codes.
- For genetic diseases resulting in organ system pathologies, use this code with appropriate keywords and relevant organ system *Pathology* codes.

CC03509 POPULATION GENETICS

Frequencies Major (12250) Minor (1580)

Applications This code retrieves studies of the genetic composition, gene frequencies and genetic drift in populations.

Examples Studies on • variations in blood group phenotypes • gene flow

Strategy Recommendations

- For mathematical models of genetic populations, use this code with appropriate keywords, the *Mathematical Biology and Statistical Methods* code CC04500 and the *Biocybernetics* code CC10515.
- For genetic studies of the evolution of populations, use this code with appropriate keywords and the *Evolution* code CC01500.

Coverage Note This code was introduced in 1972.

CC03510 SEX DIFFERENCES

Frequencies Major (25400) Minor (8840)

Applications This code retrieves studies on genetically determined sex differences and sex differentiation.

Examples Studies on • shifts of animal sex ratios • differences in enzyme activity between males and females • X-linked disorders

Strategy Recommendations

- For sex distribution studies of organ system disease patterns, use this code with appropriate keywords and relevant organ system *Pathology* codes.

PRINT INDEX SAMPLE ENTRIES

► Biosystematic Index (Sample Entry)

• Muridae

Addiction (includes Alcohol, Drugs, Smoking)

52876	52915	52929	52930	52931
52933	52934	52935	52936	

*

*

*

Genetic, Cytogenetics, Animal

47025	47026	47028	47033	47034
47035	47036	47037	47038	47040...

► Generic Index (Sample Entry)

MUS-MUSCULUS

BEHAV ANIMAL....43219

~~GENET ANIMAL....47026~~

PARAST GEN.....51226

PARAST GEN.....51250

ECOL ANIMAL.....45740

MED CLIN BAC....48784

PARASIT GEN.....51236

MUS-MUSCULUS-DOMESTICUS

BEHAV ANIMAL....43215

~~GENET ANIMAL....47026~~

reports on the attempt to sequence a statistically significant number of insertions in hemB, in order to determine whether there might be a basis for future studies to determine a molecular basis of IS2 insertional specificity. The results indicate that IS2 inserts in a non-random manner into a 240 bp segment at the 5' end of the gene (region I). Twenty-one of 24 insertions occurred in region I. Three insertions have been identified in the two middle 250 bp segments of the 975 bp gene, and none in the 3' terminal segment. A seventeen bp sequence showing 88.2% identity with a segment of IS2.221 bp from the 3' terminus has been identified in region I. Four instances of repeated insertion between the same pair of nucleotides have been observed at four different sites.

47024. Braun, Lundy*, Ryoko Mikumo, Hon Fong Mark and Stuart Lauchlan. (Dep. Pathol. Lab. Med., Box G, Brown Univ., Providence, RI 02912, USA.) *American Journal of Pathology* 143(3): 832-844, 1993. Analysis of the growth properties and physical state of the human papillomavirus type 15 genome in cell lines derived from primary cervical tumors.—We have established three cell lines from kerafinizing and nonkerafinizing cervical carcinomas with distinct growth properties in vitro and in vivo. Each cell line contained human papillomavirus type 16 DNA sequences, but the lines differed in the physical state of the viral genome present in the cells. A high copy number of episomal human papillomavirus type 16 DNA sequences was detected in the TC-140 line derived from a kerafinizing cervical cancer. This cell line had an aneuploid karyotype, did not grow in soft agar, and formed benign cyst-like nodules in nude mice, similar in morphology to well-differentiated areas of the primary tumor. Only integrated human papillomavirus type 16 sequences were detected in the TC-146A and TC-146B lines established from a nonkerafinizing large-cell cervical carcinoma. These cell lines exhibited reduced sensitivity to transforming growth factor- β 1 and produced invasive, but not progressively growing, tumors in nude mice. These cell lines should complement existing in vitro models of cervical carcinogenesis and provide useful tools for understanding the importance of virus integration in the transformation process as well as the cellular and molecular basis for tumor progression.

GENETICS AND CYTOGENETICS

See also: *Cytology and Cytochemistry • Evolution • Genetics of Bacteria and Viruses • Methods, Materials and Apparatus, General • Cytology and Cytochemistry*

See: *Specific systematic sections for cyto-taxonomic studies*

ANIMAL

See also: *Animal Production (includes Fur-Bearing Animals) • Breeds and Breeding • Developmental Biology-Embryology • Pathological, Teratology and Teratogenesis • Laboratory Animals • Metabolism • Metabolic Disorders • Poultry Production • Breeds and Breeding*

47025. Shelley, Martha E., Amjad Hossain, Paul G. McDonough and Iqbal Khan. (Reproductive Endocrinology, Sect. CJ-134, Dep. Obstet. Gynecol., Med. Coll. Ga., Augusta, GA 30912, USA.) *American Journal of Obstetrics and Gynecology* 170(3 Part 1): 1410-1415, 1994. Differential c-jun gene expression with tonically administered steroids in rat ovary and uterus.—OBJECTIVE: The purpose of this study was to evaluate the induction of the early regulatory gene c-jun in response to tonic exposure to estradiol and progesterone in rat ovary, uterus, and adrenal tissues. STUDY DESIGN: Pooled ovaries containing estradiol 17- β , progesterone, and estradiol-17 β plus progesterone were placed subcutaneously in immature female Sprague-Dawley rats (N = 24). The ovary, uterus, and the adrenal were evaluated for c-jun expression by Northern analysis at 24 and 48 hours. RESULTS: The c-jun messenger ribonucleic acid expression in the ovary and adrenal gland was inhibited with high, nonphysiologic doses of estradiol in progesterone and was induced with physiologic levels of estradiol. Physiologic levels of progesterone do not appear to influence the expression of c-jun in the ovary or adrenal gland. Uterine c-jun expression to estradiol and progesterone is generally the opposite of that observed in the ovary. CONCLUSION: These findings suggest that there is both tissue and dose specificity of c-jun gene expression in steroidogenic and steroid-responsive tissues when steroid hormones are tonically administered.

47026. Nunney, Leonard* and Ann Eileen Miller Baker. (Dep. Biol., Univ. Calif., Riverside, CA 92521, USA.) *Evolution* 47(3): 1342-1359, 1993(1994). The role of deme size, reproductive patterns, and dispersal in the dynamics of t-lethal haplotypes.—The t-lethal haplotypes (t) found in house mouse (*Mus musculus*) populations are recessive lethals favored by gametic selection whereby male heterozygotes exhibit a non-Mendelian transmission ratio of about 95% t. The expected equilibrium frequency is 0.385; however, empirical values are lower, averaging close to 0.13. We examined the hypothesis that interdemal selection is the cause of the low empirical values by using a deme-structured simulation model that included overlapping generations, a realistic breeding system, differential deme productivity, and a large total population. We found that under some conditions interdemal selection could lower t frequency below 0.13 in the face of immigration rates up to 5%. Low frequencies were correlated with effective deme size (n_e), regardless of whether n_e was changed through changing deme size (n) or through changing the proportion of breeding adults. Earlier workers showed how the first two phases of interdemal selection (random genetic differentiation and mass selection) interacted to reduce the haplotype frequency, but here we show the importance of the third phase (differential productivity of demes) once demes are linked by dispersal. The effect of this phase is not due to the (negative) covariation between deme productivity and haplotype frequency, but occurs when differential deme productivity generates a difference in t frequency between the population of juveniles recruited into their natal

deme and the population of juvenile dispersers. This difference was maximized when the average productivity of demes was low, either because few adult females bred at any one time and/or because fecundity was low. Contrary to an earlier prediction, male-biased dispersal also reduced haplotype frequency, and this probably stems from the relative excess of wild-type genotypes among dispersers compared to the deme residents. Another unexpected finding was that the randomly generated excess of heterozygotes ($F_{IS} < 0$) found in small demes favored t haplotypes; however, the effect was only seen when the more powerful influence of the third phase of interdemal selection was removed. Simulations of neutral polymorphisms showed that a deme structure giving $F_{ST} \leq 0.6$ is inconsistent with a haplotype frequency below 0.13. Based on current empirical estimates of F_{ST} (about 0.2), we concluded that immigration rates in the field are too high for interdemal selection alone to cause the observed deficit of lethal haplotypes. One factor that could combine with population structure effects is the observation that the transmission ratio is lowered to around 0.6 in litters produced from postpartum estrus (PPE). Incorporating this factor, we showed that interdemal selection could be effective in lowering the frequency of t below 0.13 when F_{ST} was above 0.43 even when migration rates were up to 10%. These results suggest that if empirical haplotype and F_{ST} estimates are accurate, then additional factors such as a lowered fitness of heterozygotes may be involved.

47027. Scribner, Kim T.*, Justin D. Congdon, Ronald K. Chesser and Michael H. Smith. (Alaska Fish and Wildlife Res. Cent., Natl. Biol. Survey, 1011 E. Tudor Road, Anchorage, Alaska 99503, USA.) *Evolution* 47(3): 1360-1373, 1993(1994). Annual differences in female reproductive success affect spatial and cohort-specific genotypic heterogeneity in painted turtles.—Long-term ecological data were used to evaluate the relative importance of movements, breeding structure, and reproductive ecological factors to the degree of spatial and age-specific variation in genetic characteristics of painted turtles (*Chrysemys picta*) on the E. S. George Reserve in southeastern Michigan. Estimates of the degree of spatial genetic structuring were based on the proportion of total genotype variance partitioned within and between subpopulations (inferred from hierarchical F-statistics based on variation at 18 protein loci) and in terms of gene correlations (co-ancestry among individuals) derived from reproductive data on full-sib families of females nesting at specific nesting areas. Little variation in allele frequency was observed among turtles from different marshes ($F_{ST} = 0.003$), though significant variation was observed among turtles from different nesting areas associated with each marsh ($F_{ST} = 0.046$). Gene correlations among individuals within nesting areas varied greatly over years (0.032-0.171; mean = 0.069) and were negatively correlated to the proportion of females that successfully nested during each year. General concordance between independent estimates of genotypic correlations (i.e., F_{ST} derived from protein electrophoretic variation vs. mean co-ancestry) suggests that allozymic data, when collected over spatial scales consistent with species behavioral characteristics and reproductive ecology, may accurately reflect the apportionment of gene diversity within and among subpopulations. The magnitude and patterning of allelic variation among nesting areas and individuals appears to be primarily a function of genetic correlations among members of full-sib families, irrespective of the degree of gene flow or female nesting-site fidelity. Comparisons of genetic characteristics among 11 cohorts (1974-1984) revealed that heterozygosity (H) and inbreeding coefficients (F) varied greatly. Cohort estimates of H and F were correlated to female nesting success and to estimates of co-ancestry for the same years. Results clearly reflect the concomitant importance of ecological factors (principally the proportion of the female population that successfully produce offspring during each year) in determining the magnitude and patterning of gene correlations within and among groups, and to the genotypic composition of offspring born during each year.

47028. Hanflie, Heidi C.* and Jeremy B. Searle. (Via Retta No. 18, 23030 Tovo S. Agata (SO), ITL.) *Evolution* 47(3): 1374-1393, 1993(1994). Extreme karyotypic variation in a *Mus musculus domesticus* hybrid zone: The tobacco mouse story revisited.—The Robertsonian fusion is a common chromosomal mutation among mammal species and is especially prevalent in the West European house mouse, *Mus musculus domesticus*. More than 40 races of the house mouse exist in Europe, including the famous "tobacco mouse" (Peschavio race) of Val Peschavio, Switzerland. Documented here is the discovery of an extreme case of karyotypic variation in the neighboring Upper Valtellina, Italy. In a 20-km stretch of the valley, 32 karyotypes were observed, including five chromosomal races and 27 hybrid types. One previously unknown race is reported, the "Mid Valtellina" race, with a diploid number of $2n = 24$ and the Robertsonian fusions Rb(1.3), Rb(4.6), Rb(5.15), Rb(7.18), Rb(8.12), Rb(9.14), Rb(11.13), and Rb(16.17). The Peschavio race ($2n = 26$), Upper Valtellina race ($2n = 24$), Lower Valtellina race ($2n = 22$) and all-acrocentric race ($2n = 40$) were also present. The races form a patchy distribution, which we term a "mottled hybrid zone." Geographical position, isolation, extinction, recolonization, and selection against hybrids are all believed to be instrumental in the origin and evolution of this complex system. Previous studies of house mice from Upper Valtellina indicated that two of the races in the valley (the Upper Valtellina and Peschavio races) may have speciated in the village of Migonico. We discuss the possibility that there may have been a reinforcement event in this village.

47029. Dickinson, W. J.*, Yifan Yang*, Kim Schinke and Michael Akam. (Dep. Biol., Univ. Utah, Salt Lake City, UT 84112, USA.) *Evolution* 47(3): 1396-1406, 1993(1994). Conservation of molecular prepatterns during the evolution of cuticle morphology in *Drosophila* larvae.—We are using patterns of cuticle specialization in *Drosophila* larvae as models to investigate the molecular, genetic, and developmental basis of morphological evolution. Members of the *virilis* species group differ markedly from one another in the distribution of hairs on the dorsal surface of first instar larvae. In particular, characteristic bands of hairs cover about 20% of each trunk segment in some species but about 70% in others. These major types do not correlate with recently proposed phylogenetic relationships, suggesting that similar phenotypes have arisen independently in different lineages. The patterns of expression of several genes that control or reflect intrasegmental patterning are indistinguishable in species with very different

CON: Nucleic Acid.Purine.Pyrimidine Studies
TAX: Enterobacteriaceae

45445. West, Stephen C. (Imperial Cancer Res. Fund, Clare Hall Lab., South Mimms, Hertfordshire EN4 3LD, UK.) *Cell* 76(1): 9-15. 1994. The processing of recombination intermediates: Mechanistic insights from studies of bacterial proteins./LITERATURE REVIEW; *ESCHERICHIA COLI*; RECOMBINANT DNA; HOLLIDAY JUNCTION BRANCH MIGRATION; RECOMBINATION ENZYME ANALYSIS; RECOMBINATION PATHWAYS; EUKARYOTIC ANALOGS

CON: Nucleic Acid.Purine.Pyrimidine Methods/Bacterial Physiology & Biochemistry/Molecular Properties & Macromolecules/Protein.Polypeptide.Amino Acid Studies/Nucleic Acid.Purine.Pyrimidine Studies/Enzyme Physiological Studies/Enzyme Chemistry & Physics

TAX: Enterobacteriaceae/Vertebrata

45446. Hall, Ruth M.* and H. W. Stokes. (CSIRO Div. Biomolecular Engineering, P.O. Box 184, North Ryde, NSW 2113, AUL.) *Genetics (Dordrecht)* 90(2-3): 115-132. 1993. Integrins: Novel DNA elements which capture genes by site-specific recombination./LITERATURE REVIEW; GRAM-NEGATIVE BACTERIA; DNA RECOMBINATION; PLASMID; TRANSPOSON; GENE INSERTION; ANTIBIOTIC RESISTANCE; MOLECULAR EVOLUTION

CON: Bacterial Physiology & Biochemistry/Enzyme Physiological Studies/Evolution/Antibacterial Chemotherapy

TAX: Bacteria

45447. Engelberg-Kulka, Hanna* and Rachel Schouler-Schwarz. (Dep. Molecular Biol., Hebrew Univ., Hadasah Med. Sch., Jerusalem 91010, ISR.) *Molecular Microbiology* 11(1): 3-8. 1994. Regulatory implications of translational frameshifting in cellular gene expression./LITERATURE REVIEW; *ESCHERICHIA COLI*; GENE REGULATION

CON: Protein.Polypeptide.Amino Acid Metabolism/Bacterial Physiology & Biochemistry/Replication, Transcription & Translation

TAX: Enterobacteriaceae

45448. Kolodner, Richard*, Sharyn D. Hall* and Cynthia Luini-Deluca. (Div. Cellular and Molecular Biol., Dana-Farber Cancer Inst., 44 Binney St., Boston, MA 02115, USA.) *Molecular Microbiology* 11(1): 23-30. 1994. Homologous pairing proteins encoded by the *Escherichia coli* *recE* and *recT* genes./LITERATURE REVIEW; *ESCHERICHIA COLI*; EXONUCLEASE VIII; ALTERNATE RECOMBINATION PATHWAY

CON: Replication, Transcription & Translation/Bacterial Physiology & Biochemistry/Protein.Polypeptide.Amino Acid Studies/Enzyme Physiological Studies

TAX: Enterobacteriaceae

GENETICS AND CYTOGENETICS

See also: Cytology and Cytochemistry • Evolution • Genetics of Bacteria and Viruses • Methods, Materials and Apparatus, General - Cytology and Cytochemistry

See: Specific systematic sections for cyto-taxonomic studies

ANIMAL

See also: Animal Production (includes Fur-Bearing Animals) - Breeds and Breeding • Developmental Biology-Embryology • Pathological, Teratology and Teratogenesis • Laboratory Animals • Metabolism • Metabolic Disorders • Poultry Production - Breeds and Breeding

45449. Mitton, Jeffery B. (Dep. Environ. Population and Organismic Biol., Univ. Colorado, Boulder, CO 80309, USA.) *Genetics (Dordrecht)* 89(1-3): 47-61. 1993. Enzyme heterozygosity, metabolism, and developmental stability./LITERATURE REVIEW; *DROSOPHILA*; *APIS MELLIFERA*; *FUNDULUS HETEROCLOTUS*; HUMAN; MONARCH BUTTERFLY; BLUE MUSSEL; SIDE-BLOTCHED LIZARD; KILLIFISH; SALMONID FISH; GUPPY; SONORAN TOPMINNOWS; HERRING; RUFOUS-COLLARED SPARROW; HOUSE SPARROW; BROWN HARE; WHITE-TAILED DEER; LACTATE DEHYDROGENASE; MORPHOLOGICAL VARIANCE; DEVELOPMENTAL HOMEOSTASIS

CON: Enzyme Physiological Studies/Comparative Biochemistry/Insect Physiology/Comparative & Experimental Mollusca/Human Genetics & Cytogenetics/Gross Anatomy/Morphogenesis

TAX: Hymenoptera/Osteichthyes/Lepidoptera/Gruiformes/Pelecaniformes/Hominidae/Hominidae/Lepidoptera/Corvidae/Diptera/Sauria

45450. Zakharov, Vladimir M. (Inst. Dev. Biol., Russian Acad. Sci., 26 Vavilov St., Moscow 117808, RUS.) *Genetics (Dordrecht)* 89(1-3): 227-234. 1993. Appearance, fixation and stabilization of environmentally induced phenotypic changes as a microevolutionary event./LITERATURE REVIEW; RESEARCH ARTICLE; *DROSOPHILA SUBOSCURA*; *LACERTA STRIGATA*; *LACERTA AGILIS BOHEMICA*; *LACERTA AGILIS LIGULA*; INTERPOPULATION PHENOTYPE DIFFERENCES CONCORDANCE; MICROPHYLOGENESIS

CON: General & Systematic Reptilia/General & Systematic Diptera/Gross Anatomy/Invertebrate Body Parts/Population Genetics/Insect Morphology/Evolution

TAX: Diptera/Sauria

45451. Watt, Ward B. (Dep. Biological Sci., Stanford Univ., Stanford, CA

94305-5020, USA.) *Genetics* 136(1): 11-16. 1994. Allazymes in evolutionary genetics: Self-imposed burden or extraordinary tool?/LITERATURE REVIEW
CON: Animal Cytology & Cytochemistry/Enzyme Physiological Studies/Evolution
TAX: Animalia

45452. Lehn, Pierre. (Univ. Rene-Descartes, CHU Paris-Ouest, Hop. Robert-Debre, Lab. de Biochim. Genetique, 48 Boulevard Serurier, 75019 Paris, FRA.) *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales* 187(4): 487-507. 1993. [In Fr. with Fr. and Engl. summ.] Current status of gene transfer into autologous hematopoietic cells./LITERATURE REVIEW; MOUSE; GENETIC ENGINEERING

CON: Lymphatic Tissue & Reticuloendothelial System/Therapy/Animal Cytology & Cytochemistry

TAX: Muridae

45453. Gilmore, Thomas D. and Patrice J. Morin. (Dep. Biol., Boston Univ., 5 Cummings St., Boston, MA 02215, USA.) *Trends in Genetics* 9(12): 427-431. 1993. The *lacB* protein: Members of a multifunctional family./LITERATURE REVIEW; ANIMAL; REL/NF-KAPPA B TRANSCRIPTION FACTOR; DNA-PROTEIN INTERACTION; MOLECULAR SEQUENCE DATA; AMINO ACID SEQUENCE

CON: Nucleic Acid.Purine.Pyrimidine Studies/Comparative & Experimental Protocols/Molecular Properties & Macromolecules/Protein.Polypeptide.Amino Acid Studies/Replication, Transcription & Translation/Animal Cytology & Cytochemistry

TAX: Animalia

45454. Werner, R. G.* and W. Noe. (Birkendort Str. 65, D-88397 Biberach/Riss, GER.) *Arzneimittel-Forschung* 43(11): 1242-1249. 1993. [In Engl. with Engl. and Ger. summ.] Mammalian cell cultures: Part II. Genetic engineering, protein glycosylation, fermentation and process control./LITERATURE REVIEW; GENE EXPRESSION

CON: Protein.Polypeptide.Amino Acid Metabolism/Animal Cytology & Cytochemistry/Replication, Transcription & Translation/Carbohydrate Metabolism/Tissue Culture

TAX: Mammalia

45455. Copertino, Donald W.* and Richard B. Hallick. (Dep. Biochemistry, Univ. Ariz., Tucson, AZ 85721, USA.) *Trends in Biochemical Sciences* 18(12): 467-471. 1993. Group II and group III introns of tRNAs: Potential relationships with nuclear pre-mRNA introns./LITERATURE REVIEW; *EUGLENA GRACILIS*; MESSENGER RNA; MOLECULAR SEQUENCE DATA; NUCLEOTIDE SEQUENCE

CON: Replication, Transcription & Translation/Plant Genetics & Cytogenetics

TAX: Euglenophyta/Flagellata

45456. Mginets, V. A. (Inst. Genet. Human. Acad. Med. Sci. Russ., Moscow, RUS.) *Genetics* 29(11): 1765-1781. 1993. [In Russ. with Russ. and Engl. summ.] Genetic regulation of limb development in vertebrates./LITERATURE REVIEW; MORPHOGENETIC FIELDS; SIMPLE GRADIENT MODEL; POLAR COORDINATES MODEL; CHONDROGENESIS; PRECHEDULED CELL DEATH; LIMB FORMATION; CELL RECEPTORS; TRANSCRIPTIONAL FACTORS; REGULATORY HOMEOBOX GENES

CON: Molecular Properties & Macromolecules/Bone, Connective & Adipose Anatomy/Nucleic Acid.Purine.Pyrimidine Methods/Animal Cytology & Cytochemistry/Replication, Transcription & Translation/Descriptive Embryology/Nucleic Acid.Purine.Pyrimidine Studies

TAX: Vertebrata

45457. Gvozdev, V. A. (Inst. Mol. Genet. Acad. Sci. Russ., Moscow 123182, RUS.) *Molekulyarnaya Biologiya (Moscow)* 27(6): 1205-1217. 1993. [In Russ. with Russ. and Engl. summ.] A possible biological role and diversity of heterochromatin regions of eukaryotic genome studied in *Drosophila melanogaster* (a review)/LITERATURE REVIEW; *DROSOPHILA MELANOGASTER*; TOTAL DNA CONTENT; GENE SEQUENCE ANALYSIS

CON: Molecular Properties & Macromolecules/Animal Cytology & Cytochemistry/Nucleic Acid.Purine.Pyrimidine Studies/Insect Physiology

TAX: Diptera

45458. Oshenbott, J. G., E. A. Van Papenrecht, T. M. Boyce, M. J. Healy and R. J. Russell. (CSIRO Div. Entomol., GPO Box 1700, Canberra ACT 2601, AUL.) *Genetics (Dordrecht)* 90(2-3): 219-264. 1993. Evolutionary genetics of *Drosophila obscura*./LITERATURE REVIEW; *DROSOPHILA MELANOGASTER*; EST6 CARBOXYL ESTERASE; ACETYLCHOLINESTERASE; SEQUENCE COMPARISON; INSECTICIDE RESISTANCE; GENE MAPPING

CON: Enzyme Physiological Studies/Post Control & Pesticides/Insect Physiology/Evolution

TAX: Hominidae

45459. Mekumia, J. A. (Dep. Genetics, Univ. Melbourne, Parkville, Victoria 3052, AUL.) *Genetics (Dordrecht)* 90(2-3): 227-237. 1993. Monitoring fitness and interspecific interactions: The evolution of resistance to diazinon in *Lucilia cuprina*./LITERATURE REVIEW; *LUCILLA CUPRINA*; INSECTICIDE; RECESSIVE TRAIT; NATURAL SELECTION; DEVELOPMENTAL STABILITY

CON: Post Control & Pesticides/Population Genetics/Insect Physiology/Evolution

TAX: Diptera

45460. Hope, Rory M. (Dep. Genetics, Univ. Adelaide, SA 5005, AUL.) *Genetics (Dordrecht)* 90(2-3): 165-182. 1993. Selected features of macropod genetics./LITERATURE REVIEW; *MACROPUS EUGENI*; *MONDELPHIS DOMESTICA*; *SMINTHOPSIS CRASSICAUDA*; CYTOGENETICS; KARYOTYPIC CONSERVATION; GENETIC LINKAGE; RECOMBINATION FREQUENCY; SPECIES COMPARISON

BIOSIS PREVIEWS SAMPLE RECORD

11212406 BIOSIS Number: 97412406

In vitro and in vivo genotoxicity evaluation of hormonal drugs: II. Dexamethasone

Singh H; Singh J R; Dhillon V S; Bali D; Paul H

155 Kabir Park, P.O. Khalsa Coll., Amritsar 143 001, IND

Mutation Research 308 (1). 1994. 89-97.

Full Journal Title: Mutation Research

ISSN: 0027-5107

Language: ENGLISH

Print Number: Biological Abstracts Vol. 098 Iss. 006 Ref. 083218

Genotoxicity evaluation of a widely used glucocorticoid medicine, dexamethasone, was undertaken using in vitro and in vivo assays. Analyses of chromosomal aberrations, sister-chromatid exchanges (SCEs) in human lymphocytes and micronuclei and SCEs in mouse bone marrow showed the drug to be capable of attacking the genetic material. However, the Ames/Salmonella assay, both with and without S9 mix, did not show any increase in His⁺ revertants.

Descriptors/Keywords: RESEARCH ARTICLE; MOUSE; HUMAN; LYMPHOCYTE; MICRONUCLEUS; BONE MARROW; PHARMACEUTICAL TOXICITY; CHROMOSOMAL ABERRATION; SISTER-CHROMATID EXCHANGE

Concept Codes:

- *02506 Cytology and Cytochemistry-Animal
- *02508 Cytology and Cytochemistry-Human
- *03506 Genetics and Cytogenetics-Animal
- *03508 Genetics and Cytogenetics-Human
- *15004 Blood, Blood-Forming Organs and Body Fluids-Blood Cell Studies
- *15008 Blood, Blood-Forming Organs and Body Fluids-Lymphatic Tissue and Reticuloendothelial System
- *22002 Pharmacology-General
- *22005 Pharmacology-Clinical Pharmacology (1972-)
- *22016 Pharmacology-Endocrine System
- *22504 Toxicology-Pharmacological Toxicology (1972-)
- 10060 Biochemical Studies-General

Biosystematic Codes:

86215 Hominidae

86375 Muridae

Super Taxa:

Animals; Chordates; Vertebrates; Mammals; Primates; Humans; Nonhuman Vertebrates; Nonhuman Mammals; Rodents

CONCEPT CODE SEARCHING ELECTRONIC PRODUCTS

- ▶ **Code Searching using Single Code**

Example: CC = 03508

- ▶ **Code Searching by Major level (primary and secondary)**

Example: CC = 03508/maj

- ▶ **Code Searching by Upposting**

Example: CC = 035 (Retrieves CC03502-CC03510)

- ▶ **Code Name Searching**

Example: CN = "Genetic and Cytogenetics-Human"

SEARCHER ATTITUDES ON CONCEPT CODES

► End User

- First choice tends to be natural language
- Codes are not intuitive. Must "look up code" to use properly
- Code descriptions are only available in printed Search Guide. No on-line reference for these codes exist
- More likely to search Concept Code Names

► Experienced Searchers

- Once exposed to Concept Codes, users tend to like them
- Use Concept Codes to limit subject area and improve precision
- Often use limit to major codes to refine search
- Most likely to use upposting of codes
- Codes can be keyed quickly
- Code searching can reduce on-line time and costs
- Avoids the need to create extensive lists of synonyms to isolate these broader categories

► Available Training and User Aids Play an Important Part in Shaping User Attitudes

SUPER TAXONOMIC CATEGORIES

- ▶ **Area where name based searching was well received by our users**
- ▶ **Searchable term assigned to specific Biosystematic Code ranges**

***Example:* BC = Birds
(Equivalent to searching 37 unique BC Codes)**

- ▶ **Used to retrieve references on taxonomically related organisms**
- ▶ **Higher level terms are intuitive and "user friendly"**

***Example:* Plants, Animals, Humans**

- ▶ **Reduces search time and costs**
- ▶ **Avoids the need to key in large groups of BC's**
- ▶ **Automatically handles historical changes**

SUPER TAXONOMIC GROUPS

The following is a list of Super Taxonomic Groups which BIOSIS has made available to the online search systems offering BIOSIS Previews. This chart lists all of the Super Taxonomic Groups and indicates the Biosystematic Code ranges they cover. Super Taxonomic Groups can be searched on BRS, Data-Star, DIALOG, DIMDI, ESA-IRS, OCLC and STN online search systems as indicated. An example of the correct format for entering this information is also given. Note: DIMDI excludes the Biosystematic Code for humans (BC86215 - Hominidae) from selected Super Taxonomic Groups (see NOTES column in chart for groups to which this condition applies).

BRS (PLANTS.ST.)	DATA-STAR (PLANTS#)	DIALOG (BC-PLANTS)	DIMDI (PPS-PLANTS)	ESAIRS (BC-PLANTS)	OCLC (BC-PLANTS)	STN (PLANTS/BC)
---------------------	------------------------	-----------------------	-----------------------	-----------------------	---------------------	--------------------

BIOSYSTEMATIC CODE RANGE	SUPER TAXONOMIC GROUP	NOTES
BC01000-BC09931 or BC13000-BC15900 or BC35000-BC35500	MICROORGANISMS	
BC02000-BC03600	VIRUSES	
BC04000-BC09931	BACTERIA	
BC05000-BC09302	EUBACTERIA	Term introduced in 1992.
BC09400-BC09931	ARCHAEOBACTERIA	Term introduced in 1992.
BC09200-BC09250	CYANOBACTERIA	Term introduced in 1992.
BC11000-BC26955	PLANTS	
BC12000-BC21800	NONVASCULAR PLANTS	
BC13000-BC14700	ALGAE	
BC15000-BC15900	FUNGI	
BC21000-BC21800	BRYOPHYTES	
BC22000-BC26955	VASCULAR PLANTS	
BC23000-BC23400	PTERIDOPHYTES	
BC25000-BC26955	SPERMATOPHYTES	
BC25100-BC25105	GYMNOSPERMS	
BC25200-BC26955	ANGIOSPERMS	
BC25202-BC25470	MONOCOTS	
BC25500-BC26955	DICOTS	
BC33000-BC86470	ANIMALS	Use "Animal" on DIMDI. DIMDI does not include BC86215 (HOMINIDAE).
BC34000-BC83500 or BC85100-BC85104	INVERTEBRATES	
BC35000-BC35500	PROTOZOANS	
BC44000 or BC45000-BC45300 or BC51000-BC51600	HELMINTHS	
BC45000-BC45300	PLATYHELMINTHS	
BC51000-BC51600	ASCHELMINTHS	
BC61000-BC61700	MOLLUSKS	
BC65000-BC65500	ANNELIDS	
BC75000-BC75406	ARTHROPODS	

BIOSYSTEMATIC CODE RANGE	SUPER TAXONOMIC GROUP	NOTES
BC75100-BC75114	CRUSTACEANS	
BC75200-BC75208	MYRIAPODS	
BC75300-BC75354	INSECTS	
BC75400-BC75406	CHELICERATES	
BC83000-BC83500	ECHINODERMS	
BC85000-BC86470	CHORDATES	
BC85100-BC85104	PROTOCHORDATES	
BC85150-BC86470	VERTEBRATES	DIMDI does not include BC86215 (HOMINIDAE).
BC85150-BC86210 or BC86220-BC86470	NONHUMAN VERTEBRATES	Use "Vertebrates" on DIMDI.
BC85200-BC85208	FISH	
BC85300-BC85306	AMPHIBIANS	
BC85400-BC85410	REPTILES	
BC85500-BC85570	BIRDS	
BC85700-BC86470	MAMMALS	DIMDI does not include BC86215 (HOMINIDAE).
BC87500-BC86210 or BC86220-BC86470	NONHUMAN MAMMALS	Use "Mammals" on DIMDI.
BC85705-BC85750	ARTIODACTYLS	
BC85760-BC85795	CARNIVORES	
BC85800-BC85840	CETACEANS	
BC85845-BC85930	BATS	
BC85940-BC85945	DERMOPTERANS	
BC85955-BC85970	EDENTATA	
BC85980-BC85985	HYRACOIDS	
BC85990-BC86030	INSECTIVORES	
BC86035-BC86045	LAGOMORPHS	
BC86055-BC86100	MARSUPIALS	
BC86105-BC86115	MONOTREMES	
BC86140-BC86155	PERISSODACTYLS	
BC86160-BC86165	PANGOLINS	
BC86170-BC86185	PINNIPEDS	
BC86190-BC86245	PRIMATES	DIMDI does not include BC86215 (HOMINIDAE).
BC86190-BC86210 or BC86220-BC86245	NONHUMAN PRIMATES	Use "Primates" on DIMDI.
BC86215	HUMANS	Use "Human" on DIMDI, ESA-IRS, and STN.
BC86250-BC86255	ELEPHANTS	
BC86265-BC86425	RODENTS	
BC86430-BC86440	SIRENIANS	
BC86465-BC86470	AARDVARKS	

BIOSIS EXPERIENCES WITH CATEGORY INDEXING

► Pros

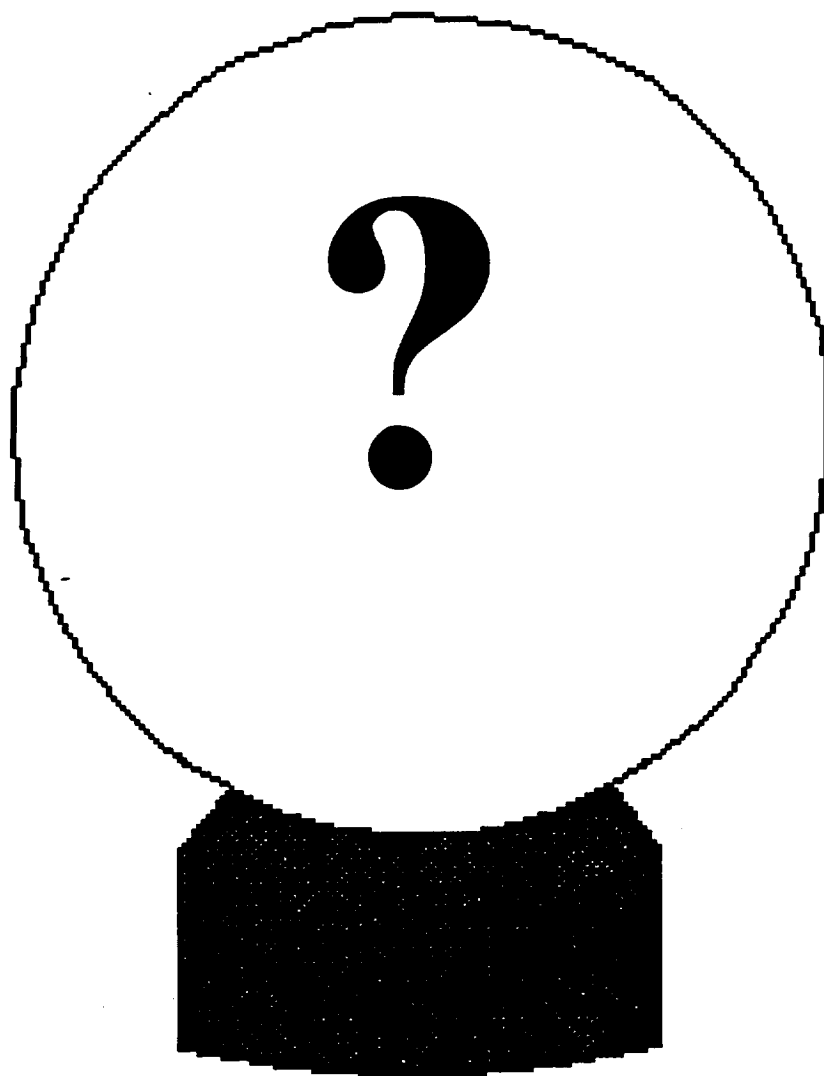
- Allows access to broad Subject Concepts
- Allows refinement of profiles based on subject matter and major/minor concepts
- Avoids need to develop lengthy lists of synonyms
- Reduces online time and costs
- Allows rapid indexing of multiple areas covered in a document
- When used with other indexing components, can improve precision of search

BIOSIS EXPERIENCES WITH CATEGORY INDEXING

► Cons

- Steep training curve to learn all the concept codes and their uses
- New areas of research and topics on the fringe of coverage may not fit into present scheme
- Alone, Concept level indexing is usually too broad for retrieval
- Maintaining and updating concept level indexing is difficult. Code based systems (with codes that go out to product and allow truncation) present additional challenges.
- If updates are performed, backfile compatibility can become an issue. For total file compatibility, reloads may be necessary
- Code assignment and use is subject to trainers and trainees background

CATEGORY INDEXING



FUTURE PLANS

CATEGORY INDEXING POINTS FOR CONSIDERATION

- 1. Your data's scope will decide how extensive your categories need to be. Use categories pertinent to the subjects found in your database.**
- 2. Keep category set to a manageable size and at same level. Avoid overlap between categories.**
 - A. If too broad**
 - If too many items are indexed to one term it becomes useless**
 - Depending on database size, searching on megacategories could take time**
 - B. If too narrow**
 - Category list becomes unmanageable**
 - Can require lots of indexing for each record.**
- 3. Develop clear editorial guidelines for the assignment of categories. When possible include indexing tools (online help, printed manuals, data validity checks) to help maintain consistency.**

CATEGORY INDEXING POINTS FOR CONSIDERATION (Cont.)

- 4. Allow some mechanism to record areas not covered by current system. This can provide new areas for expansion of categories list.**
- 5. Know the needs of your customers (both external and internal)**
- 6. Consider user support tools and systems**
- 7. Consider need to revise and update system in initial stages. This includes adding new categories and deleting unused categories.**
- 8. Decide how important backfile compatibility is to your efforts. Will mapping old system to new system be needed?**

APPENDIX

CONCEPT CODE DIRECTORY

CONCEPT CODE DIRECTORY

INTRODUCTION

Concept Codes are five-digit codes representing broad subject areas in the life sciences. Currently, over 500 Concept Codes are used in BIOSIS Previews. Throughout this search guide, the letters "CC" precede each Concept Code, allowing you to easily distinguish them from the Biosystematic Codes which are preceded by "BC." Because the Concept Codes are an important part of BIOSIS' controlled indexing and are powerful searching tools, you should always consider their use when developing search strategies. Using Concept Codes may eliminate the need to develop long lists of synonyms, saving you both time and money. Once you have selected candidate Concept Codes for your search, consult the Scope Notes section to determine whether the codes chosen are appropriate.

The Concept Code Directory lists all Concept Codes in numerical order and can be used as a quick reference to locate Concept Codes and their headings. It also can be used to identify the broad subject areas indexed in the database. The directory has four components:

Broad Concept Heading: The Concept Code Directory lists Concept Codes under subject groups called Broad Concept Headings. Use Broad Concept Headings to locate and review related codes and headings.

Concept Code: This column lists in numerical order the five-digit Concept Codes used in BIOSIS Previews.

Concept Heading: This column contains the specific Concept Heading, or title, of each Concept Code.

Notes: The Notes column provides additional useful information about the Concept Codes. For Concept Codes that have been introduced or deleted since 1969, the year of the change is listed here. If the application of a Concept Code is restricted to specific indexing levels, the restriction is stated here. Other notes indicate changes in scope for specific Concept Codes. If there is no information in the Notes column for a Concept Code, the Concept Code has been in use since 1969, has not undergone major revisions in scope, and can be applied at any indexing level.

For retrieval purposes, BIOSIS assigns Concept Codes at one of two indexing levels – MAJOR or MINOR. Most documents have multiple Concept Codes assigned at these levels. Concept Codes assigned at the MAJOR level reflect the main topics discussed in the source document. MINOR level codes are used to index ideas or concepts of considerably less emphasis.

On most search systems, you can restrict Concept Codes to the MAJOR level to ensure high relevancy. For more comprehensive retrieval, you may wish to search Concept Codes at both MAJOR and MINOR levels. Before using a Concept Code in your search strategy, you should refer to the Search System Information section of this guide or your search system documentation to determine the correct format for entering Concept Codes.

CONCEPT CODE DIRECTORY

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
GENERAL	CC00502	Philosophy	Minor level only (1980-)
	CC00504	Taxonomy, Nomenclature and Terminology	
	CC00506	Explorations, Expeditions, etc.	
	CC00508	Institutions, Administration and Legislation	
	CC00510	Museums, Botanical and Zoological Gardens, Aquaria, etc.	
	CC00512	Conservation, Resource Management	
	CC00514	Textbooks, General; Education; Audio-Visual Aids	
	CC00516	Nature Study and Materials	
	CC00518	Nontechnical Literature	
	CC00520	Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals	
	CC00522	History and Archaeology	Discontinued in 1980
	CC00524	Biographies	
	CC00526	Bibliography	
	CC00528	New Journals	
	CC00530	Information, Documentation, Retrieval and Computer Applications	
	CC00531	Forensic Science	
	CC00532	General Biology - Miscellaneous	
METHODS, MATERIALS AND APPARATUS, GENERAL	CC01004	Laboratory Methods	
	CC01006	Laboratory Apparatus	
	CC01008	Field Methods	
	CC01010	Field Apparatus	
	CC01012	Photography	
	CC01014	Apparatus - Miscellaneous	
MICROSCOPY TECHNIQUES	CC01052	General and Special Techniques	
	CC01054	Cytology and Cytochemistry	
	CC01056	Histology and Histochemistry	
	CC01058	Electron Microscopy	
EVOLUTION	CC01500	Evolution	
CYTOLOGY AND CYTOCHEMISTRY	CC02502	General	
	CC02504	Plant	
	CC02506	Animal	
	CC02508	Human	
GENETICS AND CYTOGENETICS	CC03502	General	Introduced in 1972
	CC03504	Plant	
	CC03506	Animal	
	CC03508	Human	
	CC03509	Population Genetics	
	CC03510	Sex Differences	
MATHEMATICAL BIOLOGY AND STATISTICAL METHODS	CC04500	Mathematical Biology and Statistical Methods	
PHYSICAL ANTHROPOLOGY; ETHNOBIOLOGY	CC05000	Physical Anthropology; Ethnobiology	
SOCIAL BIOLOGY; HUMAN ECOLOGY	CC05500	Social Biology; Human Ecology	

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
AEROSPACE AND UNDERWATER BIOLOGICAL EFFECTS	CC06002 CC06004 CC06006 CC06008 CC06010 CC06012	General; Methods Ecology and Psychology Physiology and Medicine Engineering and Instrumentation Exobiology Space Radiation	
SUBTERRANEAN BIORESEARCH	CC06400	Subterranean Bioresearch	Introduced in 1972
RADIATION BIOLOGY	CC06502 CC06504 CC06506	General Radiation and Isotope Techniques Radiation Effects and Protective Measures	
BEHAVIORAL BIOLOGY	CC07002 CC07003 CC07004 CC07005 CC07006	General and Comparative Behavior Animal Behavior Human Behavior Conditioning Animal Communication	Introduced in 1971
CIRCADIAN RHYTHMS AND OTHER PERIODIC CYCLES	CC07200	Circadian Rhythms and Other Periodic Cycles	
ECOLOGY; ENVIRONMENTAL BIOLOGY	CC07502 CC07504 CC07506 CC07508 CC07510 CC07512 CC07514 CC07516 CC07517 CC07518	General; Methods Bioclimatology and Biometeorology Plant Animal Oceanography and Limnology Oceanography Limnology Wildlife Management – Aquatic Water Research and Fishery Biology Wildlife Management – Terrestrial	Discontinued in 1985; minor level only
CLINICAL BIOCHEMISTRY; GENERAL METHODS AND APPLICATIONS	CC10006	Clinical Biochemistry; General Methods and Applications	
COMPARATIVE BIOCHEMISTRY, GENERAL	CC10010	Comparative Biochemistry, General	
BIOCHEMISTRY – PHYSIOLOGICAL WATER STUDIES	CC10011	Biochemistry – Physiological Water Studies	Introduced in 1970
BIOCHEMISTRY – GASES	CC10012	Biochemistry – Gases	Introduced in 1970
BIOCHEMICAL METHODS	CC10050 CC10052 CC10053 CC10054 CC10055 CC10056 CC10057	General Nucleic Acids, Purines and Pyrimidines Vitamins Proteins, Peptides and Amino Acids Porphyrins and Bile Pigments Lipids Sterols and Steroids	

CONCEPT CODE DIRECTORY

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
BIOCHEMICAL METHODS (Cont.)	CC10058	Carbohydrates	
	CC10059	Minerals	
BIOCHEMICAL STUDIES	CC10060	General	
	CC10062	Nucleic Acids, Purines and Pyrimidines	
	CC10063	Vitamins	
	CC10064	Proteins, Peptides and Amino Acids	
	CC10065	Porphyrins and Bile Pigments	
	CC10066	Lipids	
	CC10067	Sterols and Steroids	
	CC10068	Carbohydrates	
	CC10069	Minerals	
REPLICATION, TRANSCRIPTION, TRANSLATION	CC10300	Replication, Transcription, Translation	Introduced in 1970
BIOPHYSICS	CC10502	General Biophysical Studies	
	CC10504	General Biophysical Techniques	
	CC10506	Molecular Properties and Macromolecules	
	CC10508	Membrane Phenomena	
	CC10510	Bioenergetics: Electron Transport and Oxidative Phosphorylation	
	CC10511	Bioengineering	Introduced in 1972
	CC10515	Biocybernetics	
EXTERNAL EFFECTS	CC10502	General	
	CC10604	Light and Darkness	
	CC10606	Pressure	
	CC10608	Sonics; Ultrasonics	
	CC10610	Electric, Magnetic and Gravitational Phenomena	Introduced in 1970
	CC10612	Physical and Mechanical Effects	Introduced in 1971
	CC10614	Temperature as a Primary Variable	Introduced in 1971
	CC10616	Temperature as a Primary Variable – Cold	Introduced in 1971
	CC10618	Temperature as a Primary Variable – Hot	Introduced in 1972
	CC10620	Humidity	
ENZYMES	CC10802	General and Comparative Studies; Coenzymes	
	CC10804	Methods	
	CC10806	Chemical and Physical	
	CC10808	Physiological Studies	
ANATOMY AND HISTOLOGY, GENERAL AND COMPARATIVE	CC11102	Gross Anatomy	Introduced in 1971
	CC11103	Comparative Anatomy	
	CC11104	Experimental Anatomy	
	CC11105	Surgery	
	CC11106	Radiologic Anatomy	
	CC11107	Regeneration and Transplantation	Introduced in 1971
	CC11108	Microscopic and Ultramicroscopic Anatomy	
CHORDATE BODY REGIONS	CC11302	General	Introduced in 1970
	CC11304	Head	Introduced in 1970
	CC11306	Facial	Introduced in 1970
	CC11308	Neck	Introduced in 1970
	CC11309	Shoulder	Introduced in 1972
	CC11310	Back and Buttocks	Introduced in 1970
	CC11311	Lumbar	Introduced in 1970
	CC11312	Thorax	Introduced in 1970
	CC11314	Abdomen	Introduced in 1970
	CC11316	Pelvis	Introduced in 1970
	CC11318	Extremities	Introduced in 1970

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
PHYSIOLOGY, GENERAL AND MISCELLANEOUS	CC12002	General	Introduced in 1970
	CC12003	Comparative	
	CC12004	Instrumentation	
	CC12006	Methods	Introduced in 1970
	CC12008	Stress	Introduced in 1970
	CC12010	Exercise and Physical Therapy	
MOVEMENT	CC12100	Movement	Introduced in 1971
PATHOLOGY, GENERAL AND MISCELLANEOUS	CC12502	General	Introduced in 1970
	CC12503	Comparative	
	CC12504	Diagnostic	
	CC12508	Inflammation and Inflammatory Disease	Introduced in 1971
	CC12510	Necrosis	
	CC12512	Therapy	
METABOLISM	CC13002	General Metabolism; Metabolic Pathways	
	CC13003	Energy and Respiratory Metabolism	
	CC13004	Carbohydrates	
	CC13006	Lipids	
	CC13008	Sterols and Steroids	
	CC13010	Minerals	
	CC13012	Proteins, Peptides and Amino Acids	
	CC13013	Porphyrins and Bile Pigments	
	CC13014	Nucleic Acids, Purines and Pyrimidines	
	CC13015	Vitamins, General	
	CC13016	Fat-Soluble Vitamins	
	CC13018	Water-Soluble Vitamins	
	CC13020	Metabolic Disorders	
NUTRITION	CC13202	General Studies, Nutritional Status and Methods	Scope changed in 1970*
	CC13203	Malnutrition; Obesity	
	CC13206	Minerals	
	CC13207	Vitamins, General	
	CC13208	Fat-Soluble Vitamins	
	CC13210	Water-Soluble Vitamins	
	CC13214	General Dietary Studies	
	CC13216	Pathogenic Diets	
	CC13218	Prophylactic and Therapeutic Diets	
	CC13220	Carbohydrates	
	CC13222	Lipids	Introduced in 1972
	CC13224	Proteins, Peptides and Amino Acids	Introduced in 1972
	CC13226	Sterols and Steroids	Introduced in 1972
FOOD TECHNOLOGY	CC13502	General; Methods	Introduced in 1970
	CC13504	Fruits, Nuts and Vegetables	
	CC13506	Milling Technology	
	CC13508	Baking Technology	
	CC13510	Cereal Chemistry	
	CC13512	Malts, Brews and Other Fermentation Products	
	CC13514	Fats and Oils	
	CC13516	Meats and Meat By-Products	
	CC13518	Dairy Products	
	CC13520	Poultry and Eggs	
	CC13522	Fish and Other Marine and Freshwater Products	
	CC13524	Sugar	
	CC13530	Evaluations of Physical and Chemical Properties	
	CC13532	Preparation, Processing and Storage	
	CC13534	Synthetic, Supplemental and Enrichment Foods	

CONCEPT CODE DIRECTORY

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
DIGESTIVE SYSTEM	CC14001 CC14002 CC14004 CC14006	General; Methods Anatomy Physiology and Biochemistry Pathology	
CARDIOVASCULAR SYSTEM	CC14501 CC14502 CC14504 CC14506 CC14508	General; Methods Anatomy Physiology and Biochemistry Heart Pathology Blood Vessel Pathology	
BLOOD, BLOOD-FORMING ORGANS AND BODY FLUIDS	CC15001 CC15002 CC15004 CC15006 CC15008 CC15010	General; Methods Blood and Lymph Studies Blood Cell Studies Blood, Lymphatic and Reticuloendothelial Pathologies Lymphatic Tissue and Reticuloendothelial System Other Body Fluids	
URINARY SYSTEM AND EXTERNAL SECRETIONS	CC15501 CC15502 CC15504 CC15506	General; Methods Anatomy Physiology and Biochemistry Pathology	
RESPIRATORY SYSTEM	CC16001 CC16002 CC16004 CC16006	General; Methods Anatomy Physiology and Biochemistry Pathology	
REPRODUCTIVE SYSTEM	CC16501 CC16502 CC16504 CC16506	General; Methods Anatomy Physiology and Biochemistry Pathology	
ENDOCRINE SYSTEM	CC17002 CC17004 CC17006 CC17008 CC17010 CC17012 CC17014 CC17016 CC17018 CC17020	General Adrenals Gonads and Placenta Pancreas Parathyroid Pineal Pituitary Thymus Thyroid Neuroendocrinology	Introduced in 1972
MUSCLE	CC17501 CC17502 CC17504 CC17506	General; Methods Anatomy Physiology and Biochemistry Pathology	
BONES, JOINTS, FASCIAE, CONNECTIVE AND ADIPOSE TISSUE	CC18001 CC18002 CC18004 CC18006	General; Methods Anatomy Physiology and Biochemistry Pathology	
COELOMIC MEMBRANES	CC18200	Coelomic Membranes; Mesenteries and Related Structures	Introduced in 1970
INTEGUMENTARY SYSTEM	CC18501 CC18502 CC18504 CC18506	General; Methods Anatomy Physiology and Biochemistry Pathology	

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
DENTAL AND ORAL BIOLOGY	CC19001 CC19002 CC19004 CC19006	General; Methods Anatomy Physiology and Biochemistry Pathology	
SENSE ORGANS, ASSOCIATED STRUCTURES AND FUNCTIONS	CC20001 CC20002 CC20004 CC20006 CC20008	General; Methods Anatomy Physiology and Biochemistry Pathology Deafness, Speech and Hearing	
NERVOUS SYSTEM	CC20501 CC20502 CC20504 CC20506	General; Methods Anatomy Physiology and Biochemistry Pathology	
PSYCHIATRY	CC21001 CC21002 CC21003 CC21004 CC21006	General; Medical Psychology and Sociology Psychopathology; Psychodynamics and Therapy Psychophysiology Addiction (Alcohol, Drugs, Smoking, etc.) Mental Retardation	
PHARMACOLOGY	CC22002 CC22003 CC22005 CC22008 CC22010 CC22012 CC22014 CC22016 CC22018 CC22020 CC22022 CC22024 CC22026 CC22028 CC22030 CC22031 CC22032	General Drug Metabolism; Metabolic Stimulators Clinical Pharmacology Blood and Hematopoietic Agents Cardiovascular System Connective Tissue, Bone and Collagen-Acting Drugs Digestive System Endocrine System Immunological Processes and Allergy Integumentary System, Dental and Oral Biology Muscle System Neuropharmacology Psychopharmacology Reproductive System; Implantation Studies Respiratory System Sense Organs, Associated Structures and Functions Urinary System	Introduced in 1972
ROUTES OF IMMUNIZATION, INFECTION AND THERAPY	CC22100	Routes of Immunization, Infection and Therapy	
TOXICOLOGY	CC22501 CC22502 CC22504 CC22505 CC22506 CC22508	General; Methods and Experimental Foods, Food Residues, Additives and Preservatives Pharmacological Toxicology Antidotes and Preventative Toxicology Environmental and Industrial Toxicology Veterinary Toxicology	Introduced in 1972 Introduced in 1972 Scope changed in 1972*
TEMPERATURE: ITS MEASUREMENT, EFFECTS AND REGULATION	CC23001 CC23004 CC23005 CC23006 CC23007 CC23008 CC23010 CC23012	General Measurement and Methods Cryobiology Thermotherapy Hypothermia, Hyperthermia Thermopathology Thermorhythms Thermoadaptation Thermoregulation	Introduced in 1972 Introduced in 1971

CONCEPT CODE DIRECTORY

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
NEOPLASMS AND NEOPLASTIC AGENTS	CC24001	Diagnostic Methods	
	CC24002	General	
	CC24003	Immunology	
	CC24004	Pathology; Clinical Aspects; Systemic Effects	
	CC24005	Neoplastic Cell Lines	
	CC24006	Biochemistry	
	CC24007	Carcinogens and Carcinogenesis	
	CC24008	Therapeutic Agents; Therapy	
	CC24010	Blood and Reticuloendothelial Neoplasms	
GERONTOLOGY	CC24500	Gerontology	
PEDIATRICS	CC25000	Pediatrics	
DEVELOPMENTAL BIOLOGY - EMBRYOLOGY	CC25502	General and Descriptive	
	CC25503	Pathological	
	CC25504	Experimental	
	CC25508	Morphogenesis, General	
	CC25552	Descriptive Teratology and Teratogenesis	
	CC25554	Experimental Teratology and Teratogenesis	
ANIMAL PRODUCTION	CC26502	General; Methods	
	CC26504	Feeds and Feeding	
	CC26506	Breeds and Breeding	
POULTRY PRODUCTION	CC27002	General; Methods	
	CC27004	Feeds and Feeding	
	CC27006	Breeds and Breeding	
LABORATORY ANIMALS	CC28000	Laboratory Animals	Discontinued in 1970 Introduced in 1970 Introduced in 1970
	CC28002	General	
	CC28004	Gnotobiology	
MICROORGANISMS, GENERAL	CC29500	Microorganisms, General	
BACTERIOLOGY, GENERAL AND SYSTEMATIC	CC30000	Bacteriology, General and Systematic	
MORPHOLOGY AND CYTOLOGY OF BACTERIA	CC30500	Morphology and Cytology of Bacteria	
PHYSIOLOGY AND BIOCHEMISTRY OF BACTERIA	CC31000	Physiology and Biochemistry of Bacteria	
GENETICS OF BACTERIA AND VIRUSES	CC31500	Genetics of Bacteria and Viruses	
MICROBIOLOGICAL APPARATUS, METHODS AND MEDIA	CC32000	Microbiological Apparatus, Methods and Media	
MICROBIOLOGICAL ULTRASTRUCTURE	CC32300	Microbiological Ultrastructure	Introduced in 1972

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
TISSUE CULTURE, APPARATUS, METHODS AND MEDIA	CC32500	Tissue Culture, Apparatus, Methods and Media	
IN VITRO STUDIES, CELLULAR AND SUBCELLULAR	CC32600	In Vitro Studies, Cellular and Subcellular	
VIROLOGY	CC33502 CC33504 CC33506 CC33508	General; Methods Bacteriophage Animal Host Viruses Plant Host Viruses	
IMMUNOLOGY AND IMMUNOCHEMISTRY	CC34502 CC34504 CC34506 CC34508	General; Methods Bacterial, Viral and Fungal Immunohematology, Blood Groups Immunopathology, Tissue Immunology	
IMMUNOLOGY, PARASITOLOGICAL	CC35000	Immunology, Parasitological	
ALLERGY	CC35500	Allergy	
MEDICAL AND CLINICAL MICROBIOLOGY	CC36001 CC36002 CC36006 CC36007 CC36008 CC36502 CC36504 CC36506 CC36508	General; Methods and Techniques Bacteriology Virology Phycology Mycology Clinical Microbiological Methods; General Serodiagnosis Skin Tests Clinical Microbiological Methods; Techniques	Scope changed in 1971* Introduced in 1970 Discontinued in 1971 Discontinued in 1971
PUBLIC HEALTH	CC37001 CC37003 CC37006 CC37008 CC37010 CC37012	General and Miscellaneous Epizootiology Public Health Laboratory Methods Disinfection and Vector Control; Pesticides Public Health Administration and Statistics Health Services and Medical Care	
PUBLIC HEALTH: ENVIRONMENTAL HEALTH	CC37013 CC37014 CC37015 CC37017 CC37019	Occupational Health Sewage Disposal and Sanitary Measures Air, Water and Soil Pollution Radiation Health Miscellaneous	Scope changed in 1970*
PUBLIC HEALTH: EPIDEMIOLOGY	CC37052 CC37054 CC37056	Communicable Diseases Organic Diseases and Neoplasms Miscellaneous	
PUBLIC HEALTH: DISEASE VECTORS	CC37057 CC37058 CC37060	General Animate Inanimate	
PUBLIC HEALTH: MICROBIOLOGY	CC37400	Public Health Microbiology	
VETERINARY SCIENCE	CC38002 CC38004 CC38006	General; Methods Pathology Microbiology	

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
CHEMOTHERAPY	CC38502 CC38504 CC38506 CC38508 CC38510	General; Methods; Metabolism Antibacterial Agents Antiviral Agents Antifungal Agents Antiparasitic Agents	
FOOD AND INDUSTRIAL MICROBIOLOGY	CC39002 CC39003 CC39004 CC39006 CC39007 CC39008	Food and Beverage Spoilage and Contamination Food and Beverage Fermentation Antibiotics, Biologics, Other Agents Biodegradation and Biodegradation Biosynthesis, Bioassay and Fermentation General and Miscellaneous	Introduced in 1970 Scope changed in 1972*
DISINFECTION, DISINFECTANTS AND STERILIZATION	CC39500	Disinfection, Disinfectants and Sterilization	Scope changed in 1971*
SOIL MICROBIOLOGY	CC40000	Soil Microbiology	
PALEOBOTANY	CC50000	Paleobotany	
PALYNOLOGY	CC50100	Palynology	
BOTANY, GENERAL AND SYSTEMATIC	CC50501 CC50503 CC50504 CC50506 CC50510 CC50511 CC50512 CC50513 CC50514 CC50516 CC50518 CC50522 CC50524 CC50526 CC50528 CC50530	Cryptogamae Thallophyta Algae Fungi Lichenes Embryophyta Bryophyta Tracheophyta Pteridophyta Spermatophyta Gymnospermae Angiospermae Monocotyledones Dicotyledones Floristics and Distribution General and Miscellaneous	Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only Major level only
MORPHOLOGY, ANATOMY AND EMBRYOLOGY OF PLANTS	CC51000	Morphology, Anatomy and Embryology of Plants	
PLANT PHYSIOLOGY, BIOCHEMISTRY AND BIOPHYSICS	CC51502 CC51503 CC51504 CC51506 CC51508 CC51510 CC51512 CC51514 CC51516 CC51517 CC51518 CC51519 CC51520 CC51522 CC51523	Water Relations Temperature Nutrition Photosynthesis Respiration, Fermentation Growth Differentiation Reproduction Growth Substances Light and Radiation Effects Electric, Magnetic and Gravitational Phenomena Enzymes Metabolism Translocation, Accumulation Chemical Constituents Bioluminescence	

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
PLANT PHYSIOLOGY, BIOCHEMISTRY AND BIOPHYSICS (Cont.)	CC51524 CC51526	Apparatus and Methods General and Miscellaneous	
ECONOMIC BOTANY, GENERAL	CC52000	Economic Botany, General	
AGRONOMY	CC52502 CC52504 CC52506 CC52508 CC52510 CC52512 CC52514 CC52518	General, Miscellaneous and Mixed Crops Grain Crops Forage Crops and Fodder Fiber Crops Sugar Crops Tobacco Crops Oil Crops Weed Control	
SOIL SCIENCE	CC52801 CC52802 CC52803 CC52804 CC52805 CC52806 CC52807	General; Methods Soil Science, General Genesis, Morphology, Classification, Geology Soil Science, Inorganic Physics and Chemistry Soil Science, Organic Fertility and Applied Studies	Introduced in 1970 Discontinued in 1970 Discontinued in 1970 Introduced in 1970 Discontinued in 1970 Introduced in 1970
HORTICULTURE	CC53002 CC53004 CC53006 CC53008 CC53010 CC53012	Temperate Zone Fruits and Nuts Tropical and Subtropical Fruits and Nuts; Plantation Crops Small Fruits Vegetables Flowers and Ornamentals General; Miscellaneous and Mixed Crops	Scope changed in 1971* Scope changed in 1971*
FORESTRY AND FOREST PRODUCTS	CC53500	Forestry and Forest Products	
PHARMACOGNOSY AND PHARMACEUTICAL BOTANY	CC54000	Pharmacognosy and Pharmaceutical Botany	
PHYTOPATHOLOGY	CC54502 CC54504 CC54506 CC54507 CC54508 CC54510 CC54512 CC54514 CC54516 CC54518	Diseases Caused by Fungi Diseases Caused by Bacteria Diseases Caused by Phanerogams Diseases Caused by Algae Diseases Caused by Animal Parasites Diseases Caused by Viruses Nonparasitic Diseases Parasitism and Resistance Disease Control General and Miscellaneous	Introduced in 1971
PEST CONTROL, GENERAL; PESTICIDES; HERBICIDES	CC54600	Pest Control, General; Pesticides; Herbicides	
ECONOMIC ENTOMOLOGY	CC60002 CC60004 CC60006	General Field, Flower and Truck Crops Fruits and Nuts	Scope changed in 1971*

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
ECONOMIC ENTOMOLOGY (Cont.)	CC60008	Stored Products	
	CC60010	Trees, Ornaments and Wood Products	
	CC60012	Animal Pests	
	CC60014	Biological Control	
	CC60015	Integrated Control	
	CC60016	Chemical and Physical Control, General; Apparatus	
	CC60018	Apiculture	
	CC60020	Sericulture	
PARASITOLOGY	CC60502	General	
	CC60504	Medical	
	CC60506	Veterinary	
SANITARY ENTOMOLOGY	CC61000	Sanitary Entomology	
CHORDATA, GENERAL AND SYSTEMATIC ZOOLOGY	CC62502	General	Major level only
	CC62503	Protochordata	Major level only
	CC62510	Pisces	Major level only
	CC62512	Amphibia and Reptilia	Major level only
	CC62514	Amphibia	Major level only
	CC62516	Reptilia	Major level only
	CC62518	Aves	Major level only
	CC62520	Mammalia	Major level only
ANIMAL DISTRIBUTION	CC62800	Animal Distribution	Introduced in 1971
PALEOZOOLOGY	CC63000	Paleozoology	
INVERTEBRATA, GENERAL AND SYSTEMATIC ZOOLOGY	CC63501	General	Major level only
	CC63502	Protozoa	Major level only
	CC63504	Mesozoa	Major level only
	CC63506	Porifera	Major level only
	CC63508	Cnidaria	Major level only
	CC63509	Ctenophora	Major level only
	CC63510	Platyhelminthes	Major level only
	CC63512	Rhynchocoela	Major level only
	CC63514	Acanthocephala	Major level only
	CC63516	Aschelminthes	Major level only
	CC63518	Entoprocta	Major level only
	CC63520	Phoronidea	Major level only
	CC63522	Ectoprocta	Major level only
	CC63524	Brachiopoda	Major level only
	CC63526	Mollusca	Major level only
	CC63528	Sipunculoidea	Major level only
	CC63530	Annelida	Major level only
	CC63532	Echiuroidea	Major level only
	CC63534	Linguatulida	Major level only
	CC63536	Tardigrada	Major level only
	CC63538	Onychophora	Major level only
	CC63542	Chaetognatha	Major level only
	CC63544	Hemichordata	Major level only
	CC63546	Pogonophora	Major level only
	CC63548	Echinodermata	Major level only
	CC63552	Arthropoda - General	Major level only
	CC63554	Arthropoda - Crustacea	Major level only
	CC63556	Arthropoda - Myriapoda	Major level only
	CC63572	Insecta - General	Major level only
	CC63573	Insecta - Coleoptera	Major level only
	CC63574	Insecta - Diptera	Major level only

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
INVERTEBRATA, GENERAL AND SYSTEMATIC ZOOLOGY (Cont.)	CC63576	Fossil Insecta	Major level only
	CC63578	Insecta - Hemiptera, Heteroptera	Major level only
	CC63580	Insecta - Homoptera	Major level only
	CC63582	Insecta - Hymenoptera	Major level only
	CC63584	Insecta - Lepidoptera	Major level only
	CC63586	Insecta - Neuroptera	Major level only
	CC63588	Insecta - Orthoptera	Major level only
	CC63590	Insecta - Other Orders	Major level only
	CC63592	Chelicerata - General	Major level only
	CC63594	Chelicerata - Arachnida	Major level only
	CC63596	Chelicerata - Acarina	Major level only
	CC63597	Chelicerata - Merostomata	Major level only
	CC63598	Chelicerata - Pycnogonida	Major level only
INVERTEBRATA, COMPARATIVE AND EXPERIMENTAL MORPHOLOGY, PHYSIOLOGY AND PATHOLOGY	CC64001	General	
	CC64002	Protozoa	
	CC64004	Mesozoa	
	CC64006	Porifera	
	CC64008	Cnidaria	
	CC64009	Ctenophora	
	CC64010	Platyhelminthes	
	CC64012	Rhynchozoela	
	CC64014	Acanthocephala	
	CC64016	Aschelminthes	
	CC64018	Entoprocta	
	CC64020	Phoronidea	
	CC64022	Ectoprocta	
	CC64024	Brachiopoda	
	CC64026	Mollusca	
	CC64028	Sipunculoidea	
	CC64030	Annelida	
	CC64032	Echiuroidea	
	CC64034	Linguatulida	
	CC64036	Tardigrada	
	CC64038	Onychophora	
	CC64042	Chaetognatha	
	CC64044	Hemichordata	
	CC64046	Pogonophora	
	CC64048	Echinodermata	
	CC64052	Arthropoda - General	
	CC64054	Arthropoda - Crustacea	
	CC64056	Arthropoda - Myriapoda	
	CC64060	Arthropoda - Chelicerata	
	CC64072	Insecta - General	
	CC64074	Insecta - Morphology, Comparative	
	CC64076	Insecta - Physiology	
	CC64078	Insecta - Pathology	
INVERTEBRATE BODY REGIONS AND STRUCTURES PALEOBIOLOGY	CC64202	General	Introduced in 1971
	CC64204	Head	Introduced in 1971
	CC64206	Cephalothorax	Introduced in 1971
	CC64208	Thorax	Introduced in 1971
	CC64210	Abdomen	Introduced in 1971
	CC64212	Appendages	Introduced in 1971
	CC64214	Hard Parts	Introduced in 1971
	CC64216	Orifices, Pores and Cavities	Introduced in 1971
	CC64218	Special Organs	Introduced in 1971
PALEOBIOLOGY	CC64500	Paleobiology	

BROAD CONCEPT HEADING	CONCEPT CODE	CONCEPT HEADING	NOTES
GEOLOGICAL PERIODS	CC64702	General and Unspecified	Introduced in 1971
	CC64704	Precambrian	Introduced in 1971
	CC64705	Paleozoic, General	Introduced in 1972
	CC64706	Cambrian	Introduced in 1971
	CC64708	Ordovician	Introduced in 1971
	CC64710	Silurian	Introduced in 1971
	CC64712	Devonian	Introduced in 1971
	CC64714	Carboniferous – Mississippian and Pennsylvanian	Introduced in 1971
	CC64716	Permian	Introduced in 1971
	CC64717	Mesozoic, General	Introduced in 1972
	CC64718	Triassic	Introduced in 1971
	CC64720	Jurassic	Introduced in 1971
	CC64722	Cretaceous	Introduced in 1971
	CC64723	Cenozoic, General	Introduced in 1972
	CC64724	Tertiary	Introduced in 1971
	CC64726	Pleistocene	Introduced in 1971
	CC64728	Recent	Introduced in 1971

*See Master Index for details of historical coverage.

10

Action Items

Government Information Locator Service (GILS)

As part of the National Information Infrastructure, the U.S. Federal government is establishing a Government Information Locator Service (GILS) to help the public locate and access information. An Office of Management and Budget Bulletin will be published this year to provide implementing guidance specifying Federal agency responsibilities. The National Institute of Standards and Technology has also announced a Federal Information Processing Standard specifying a GILS Profile that is mandatory for Federal agencies maintaining information locators.

"Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it. In addition, it will require consistent Federal information policies designed to ensure that Federal information is made available at a fair price to as many users as possible while encouraging growth of the information industry."

*"Technology for America's Economic Growth,
A New Direction to Build Economic Strength"*

What is GILS?

GILS will identify public information resources throughout the Federal Government, describe the information available in those resources, and provide assistance in obtaining the information. It will consist of a decentralized collection of agency-based information locators and associated information services. GILS will supplement, but not necessarily supplant, other agency information dissemination mechanisms and commercial information sources.

The public will be served by GILS through intermediaries or directly. Central disseminating agencies such as the Government Printing Office and the National Technical Information Service will act as intermediaries to GILS, as will Depository Libraries, other public libraries and private sector information services. Access to GILS contents may also be accomplished through kiosks, 800 numbers, electronic mail, bulletin boards, FAX, and off-line media such as floppy disks, CD-ROM, and printed works.

While GILS will encompass a very wide range of information sources and many mechanisms for finding and delivering information, a "GILS Core" will be established as a definitive locator leading to Federal agency information resources. The GILS Core will be accessible on public networks without charge to direct users.

GILS will use network technology and the American National Standards Institute Z39.50 standard for information search and retrieval so that information can be retrieved in a variety of ways, and so that GILS direct users can ultimately gain access to many other major Federal and non-Federal information resources. GILS will also include automated linkages that facilitate electronic delivery of off-the-shelf information products, as well as guide users to data systems that support analysis and synthesis of information.

Where to find more information on the Government Information Locator Service (GILS)

The Office of Management and Budget sent a report to the Information Infrastructure Task Force (IITF) describing how GILS will be implemented. The description of GILS was prepared by Eliot Christian of the U.S. Geological Survey, working primarily with the interagency group known as the Working Group on Public Access (also known as the "Solomon's Group"). The report was reviewed by the three IITF Committees, the United States Advisory Council on the National Information Infrastructure, various Federal agencies, some non-Federal organizations and the general public through notices in both the Federal Register and the Commerce Business Daily, as well as through a public meeting held at the Department of the Interior on December 13, 1993.

The GILS report to the IITF, dated May 2, 1994, is available on the FedWorld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format).

OMB Circular A-130 and Information Locators

On June 25, 1993, the Office of Management and Budget revised Circular A-130, "Management of Federal Information Resources," to strengthen policies for managing government information (58 F.R. 36068, July 2, 1993). Circular A-130 states that availability of government information in diverse media, including electronic formats, permits the public greater flexibility in using the information, and that modern information technology presents opportunities to improve the management of government programs to provide better service to the public. It notes that the development of public electronic information networks, such as the Internet, provides an additional way for agencies to increase the diversity of information sources available to the public, and that emerging standards such as ANSI (American National Standards Institute) Z39.50 will be used increasingly to facilitate dissemination of government information in a networked environment.

Circular A-130 states that agencies shall:

- Disseminate information products on equitable and timely terms;
- Avoid establishing exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis;
- Use voluntary standards and Federal Information Processing Standards;
- Use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public;
- Take advantage of all dissemination channels, Federal and nonfederal, including State and local governments, libraries and private sector entities;
- Provide information describing how the public may gain access to agency information resources;
- Help the public locate government information maintained by or for the agency;
- Establish and maintain inventories of all agency information dissemination products;
- Develop such other aids to locating agency information dissemination products including catalogs and directories...

The Government Information Locator Service (GILS)

As part of the U. S. National Information Infrastructure, the Federal government is establishing a Government Information Locator Service (GILS) to help the public locate and access information. An Office of Management and Budget Bulletin is being published this year specifying Federal agency responsibilities and the implementation schedule. The National Institute of Standards and Technology has announced a Federal Information Processing Standard for the GILS Profile--a voluntary, international standard with mandatory application for Federal agencies establishing locators for information. As other nations also adopt such standards, the foundation will be laid for a Global Information Locator.

"Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it."

*Technology for America's Economic Growth,
A New Direction to Build Economic Strength
(United States strategic technology policy)*

What is GILS?

GILS will identify public information resources throughout the U.S. Federal Government, describe the information available in those resources, and provide assistance in obtaining the information. It will consist of a decentralized collection of agency-based information locators and associated information services. GILS will supplement, but not necessarily supplant, other agency information dissemination mechanisms and commercial information sources. GILS uses network technology and voluntary, international standards for information search and retrieval so that information can be retrieved in a variety of ways, and so that GILS direct users can find many other information resources worldwide. GILS also provides for automated linkages that facilitate electronic delivery of off-the-shelf information products, as well as guiding users to data systems that support analysis and synthesis of information.

The public will be served by GILS through intermediaries or directly. While GILS will encompass a very wide range of information sources and many mechanisms for finding and delivering information, U.S. Federal agency information resources will be identified in a common way using the specified GILS Core Elements. The set of locator records comprising the US Federal GILS Core will be accessible on public networks without charge to direct users. Central disseminating agencies such as the Government Printing Office and the National Technical Information Service will act as intermediaries to GILS, as will public libraries and commercial information services. Access to GILS contents may also be accomplished through kiosks, electronic mail, bulletin boards, and off-line media such as floppy disks, CD-ROM, and printed works.

Where to find more information on the Government Information Locator Service (GILS).

The U.S. Federal government prepared a report with public input describing how GILS will be implemented. The report, dated May 2, 1994, is available on the FedWorld electronic bulletin board (703-321-8020) or via the Internet by anonymous FTP (File Transfer Protocol) at 130.11.48.107 in the directory /pub as the file gils.doc (Microsoft Word for Windows format).

Toward a Global Information Locator

Why a Global Information Locator?

Although the global information revolution continues to make ever more vast amounts of information available, not all of the information resources are made known in a common manner. Users have extreme difficulty just in trying to find relevant materials. A Global Information Locator would have immediate practical application in international areas such as Agenda 21, global change research, environmental monitoring, coordination of humanitarian assistance, and U.N. administration.

The Global Information Infrastructure is emerging at a revolutionary period in the history of information. Technological breakthroughs have expanded radically the possibilities for electronic access. In particular, peer computer networks, e.g., Open Systems Interconnection (OSI) and the Internet, allow for a decentralized approach to information dissemination. On such networks, many different information sources can be maintained separately yet any user can choose to view them as a whole. The U.S. Government Information Locator Service (GILS) uses a decentralized network approach based on international standards that may serve as the model for a Global Information Locator.

How could a Global Information Locator be established?

The U.S. GILS initiative has several characteristics that are important for a Global Information Locator. Being decentralized and based on open standards, it fits the decentralized character of information dissemination globally. It encourages dissemination by a wide diversity of sources, both public and private, that serve the myriad public and governmental needs for information.

The open systems design assures that many different information systems can be separately developed yet be interoperable when implemented. Interoperability depends on a stable reference, known as an application profile, that is openly negotiated among implementors, documented, and made widely known. For the Global Information Locator, an International Standardized Profile could adapt the existing GILS Profile to establish common practices for identifying and describing information resources globally. The profile would state the functions and environments within which it applies, and would identify options and parameters of existing standards needed to achieve a Global Information Locator. As with the GILS Profile, the Global Information Locator profile would be compatible with the Internet as well as OSI-compliant networks, and would make use of open standards for information search and retrieval such as ISO 10162/10163.

The application profile for the Global Information Locator would not limit how information is maintained at the source nor how information is displayed to users. Alternative ways to organize and present networked information would continue to be encouraged, but participants in the Global Information Locator would support such alternatives in addition to supporting the profile.

Success of the Global Information Locator does not depend on massive government investments or sweeping international agreements. Rather, by adopting existing international information standards, it could build on the efforts of the responsible and talented people worldwide already working on information access issues. For example, software that supports direct access will be available for free from many sources worldwide and will also be embedded within commercial computer applications ranging from the very simple to those that dynamically interpret natural language, or filter search requests to sift huge amounts of information automatically.

The Government Information Locator Service (GILS)

*Report to the Information Infrastructure Task Force
May 2, 1994*

The Government Information Locator Service (GILS)

Executive Summary

In coordination with the Information Infrastructure Task Force (IITF), the Office of Management and Budget (OMB) is promoting the establishment of an agency-based Government Information Locator Service (GILS) to help the public locate and access information throughout the Federal Government. This report presents a vision of how GILS will be implemented.

Working primarily with OMB and the Locator Subgroup of the Interagency Working Group on Public Access, Eliot Christian of the U.S. Geological Survey prepared this report under the auspices of the IITF Committee on Information Policy. This vision of GILS has also received extensive review by various Federal agencies and other interested parties, including some non-Federal organizations and by the general public through notices in both the Federal Register and the Commerce Business Daily and at a public meeting held in December, 1993.

As part of the Federal role in the National Information Infrastructure, GILS will identify and describe information resources throughout the Federal government, and provide assistance in obtaining the information. It will be decentralized and will supplement other agency and commercial information dissemination mechanisms.

The public will use GILS directly or through intermediaries, such as the Government Printing Office, the National Technical Information Service, the Federal depository libraries, other public libraries, and private sector information services. Direct users will have access to a GILS Core accessible on the Internet without charge. Intermediate access may include kiosks, "800 numbers," electronic mail, bulletin boards, FAX, and off-line media such as floppy disks, CD-ROM, and printed works.

GILS will use standard network technology and the American National Standards Institute Z39.50 standard for information search and retrieval so that information can be retrieved in a variety of ways. Direct users will eventually have access to many other Federal and non-Federal information resources, linkages to data systems, and electronic delivery of information products.

Development of this report proceeded in tandem with a GILS Profile development project that produced an Implementors Agreement in the voluntary standards process. The National Institute of Standards and Technology is now establishing a Federal Information Processing Standard referencing the GILS Profile Implementors Agreement and making mandatory its application for Federal agencies establishing locators for government information.

Existing law and policy, as articulated in OMB Circular A-130, the Records Disposal Act, and the Freedom of Information Act, require agencies to create and maintain an inventory of their information systems and information dissemination products. Although compliance with these requirements varies greatly, the incremental cost of making those inventories accessible through GILS is expected to be minimal. Accordingly, participation in establishing and maintaining GILS may be accomplished as a collective effort executed within existing funds and authorities. OMB will publish in 1994 a Bulletin following on Circular A-130 that will specify agency responsibilities in GILS and set implementation schedules. A process for ongoing evaluation will also be established to evaluate the degree to which GILS meets the information needs of the public.

The Government Information Locator Service (GILS)

Introduction

Government information is fundamental to modern societies. Although individual Federal agencies may recognize their responsibility to maintain readily accessible inventories of their records and other information resources, there needs to be a collective vision across the Federal government for information dissemination to the public. The vision of a Government Information Locator Service (GILS) presented here responds to that need and places this Federal vision in the context of broader issues such as promotion of diverse information services.

GILS is emerging at a revolutionary period in the history of information processing where technological breakthroughs have radically expanded the range of feasible strategies. In particular, the realization of peer computer networks allows for a decentralized approach where many different information sources are separately maintained yet are comprehensible as a coherent whole from the unique perspective of a specific user. GILS depends on this network approach to preserve the decentralized character of Federal information dissemination and the wide diversity of sources, both public and private, that serve the public need for information access.

In contrast to a centralized design, a decentralized approach assumes that many different implementations will be separately developed yet will be fully interoperable when implemented. Achieving interoperability is only possible if a stable base of reference is documented and made widely known. In GILS, that reference base is an agreement among active implementors together with Federal representatives. Where fundamental design choices have been made in developing the implementors agreement, those choices have emphasized the use of stable but extensible standards.

The success of GILS does not depend on massive Federal investment or sweeping new directives. Rather, it adopts voluntary information standards in order to build on the efforts of the responsible, talented, and creative people throughout Government and in society already working on information access issues. GILS will use this solid base of widely accepted standards to help agencies and information services focus their initiatives and thereby make the vast range of Government information more accessible to the public.

Policy Context

The Administration's strategic technology policy document entitled "Technology for America's Economic Growth, A New Direction to Build Economic Strength" states:

Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it. In addition, it will require consistent Federal information policies designed to ensure that Federal information is made available at a fair price to as many users as possible while encouraging growth of the information industry.¹

On June 25, 1993, the Office of Management and Budget (OMB) revised Circular A-130, "Management of Federal Information Resources," to strengthen policies for managing government information (58 F.R. 36068, July 2, 1993). Circular A-130 encourages agencies to use new technologies to make government information available to the public in a timely and equitable manner via a diverse array of sources, both public and private. It states that availability of government information in diverse media, including electronic formats, permits the public greater flexibility in using the information, and that modern information technology presents opportunities to improve the management of government programs to provide better service to the public. It also notes that the development of public electronic information networks, such as the Internet, provides an additional way for agencies to increase the diversity of information sources available to the public, and that emerging standards such as ANSI (American National Standards Institute) Z39.50² will be used increasingly to facilitate dissemination of government information in a networked environment.

OMB Circular A-130 states that agencies shall:

- Disseminate information products on equitable and timely terms;
- Avoid establishing, or permitting others to establish on their behalf, exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis;
- Use voluntary standards and Federal Information Processing Standards where appropriate or required;
- Use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public;

¹ Clinton, William J. & Gore, Albert, Jr., (1993, February 22). Technology for America's Strength. A New Direction to Build Economic Strength. Washington, DC: Government Printing Office.

² National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992. Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Gaithersburg, MD: National Information Standards Organization Press.

- Take advantage of all dissemination channels, Federal and nonfederal, including State and local governments, libraries and private sector entities;
- Provide information describing how the public may gain access to agency information resources;
- Help the public locate government information maintained by or for the agency;
- Establish and maintain inventories of all agency information dissemination products;
- Develop such other aids to locating agency information dissemination products including catalogs and directories...

Because the active management of information by agencies is essential to the operation of government and to democratic principles, laws and policies assert a fundamental requirement that Federal agencies maintain readily accessible inventories of their records and other information holdings. The responsibilities of Federal agencies with regard to the management of electronic records are also growing in importance as their reliance on electronic information systems increases. To help the public locate and gain access to public information within agency inventories, the Administration has committed to promote the establishment of an agency-based Government Information Locator Service (GILS).

Working primarily with OMB and the Locator Subgroup of the Interagency Working Group on Public Access (the "Solomon's Group"), Eliot Christian of the U.S. Geological Survey (USGS) prepared this report to the Information Infrastructure Task Force describing how GILS may be implemented. Development of this report proceeded in tandem with a GILS Profile development project that produced an Implementors Agreement in the voluntary standards process. The GILS Profile project was a Cooperative Agreement between the USGS and Syracuse University, funded by the Interagency Working Group on Data Management for Global Change, with active involvement from several ANSI Z39.50 implementors representing non-government sectors.³ The National Institute of Standards and Technology (NIST) is now establishing a Federal Information Processing Standard (FIPS) referencing the GILS Profile Implementors Agreement and making mandatory its application for Federal agencies establishing locators for government information.

Existing law and policy, as articulated in OMB Circular A-130, the Records Disposal Act (Title 44 of the United States Code), and the Freedom of Information Act (FOIA), already require agencies to create and maintain an inventory of their information systems and information dissemination products. Although compliance with these requirements varies greatly, the incremental cost of making those inventories accessible through GILS is expected to be minimal. Accordingly, participation in meeting the minimum mandatory requirements for establishing and maintaining GILS may be accomplished as a collective effort within existing funds and authorities.

OMB will publish in 1994 a Bulletin following on Circular A-130 that will specify agency responsibilities in GILS and set implementation schedules. A process for ongoing evaluation will also be established to evaluate the degree to which GILS meets user information needs, including

³ McClure, Charles R., & Moen, William E. (1994). Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard. Syracuse, NY: School of Information Studies, Syracuse University.

factors such as accessibility, ease of use, suitability of descriptive language, accuracy, consistency, timeliness, and completeness of coverage.

The User Perspective

GILS must be many things to many people. It must be comprehensive, yet user friendly. It must answer specific questions, yet enable scanning a wide range of government information. It must be able to answer questions from the most inexperienced users, yet permit in-depth research as well. It must be of direct service to the public, yet not undermine the diversity of existing information sources. Private-sector information providers must be able to participate in GILS and also make their resources known and accessible.

GILS depends critically on other aspects of the emerging NII. GILS must be implemented with full recognition of individual privacy and intellectual property rights. Agencies will need to ensure that members of the public whom the agency has a responsibility to inform have a reasonable ability to access GILS and the underlying information resources and information dissemination products. Agencies participating in GILS must take care to minimize barriers to use, including equipment and software requirements, cost, and technical complexity.

The public will use GILS either directly or through intermediaries. The distinction is that direct users roam at will, but users of intermediate services take a guided tour. The following are some examples of GILS direct users and intermediaries:

- A direct user researching national health care may explore relevant issues from a variety of perspectives by accessing a wide range of GILS and non-GILS information sources.
- An educator interested in keeping up with electronic educational materials may access a few GILS sources once a month as a direct user over a dial-up connection to the Internet.
- An information service may query GILS hourly as a direct user and also act as an intermediary by constructing a value-added directory derived from GILS for sale to users who need specific products such as government economic statistics.
- A Federal agency may act as an intermediary in adding GILS access into its existing information service to provide public information referrals to sources in other agencies.

A major advantage of the networked and decentralized design of GILS is that it allows direct users to explore many different aspects of government information. Since direct users are less limited in their searching, they have more flexibility to explore the full complement of available information. For direct users, there is minimal structure across the GILS locator records and the records are interleaved with a vast diversity of other kinds of information. On the Internet, direct users have tools for interacting with people, news, and libraries in addition to GILS (Figure 1).

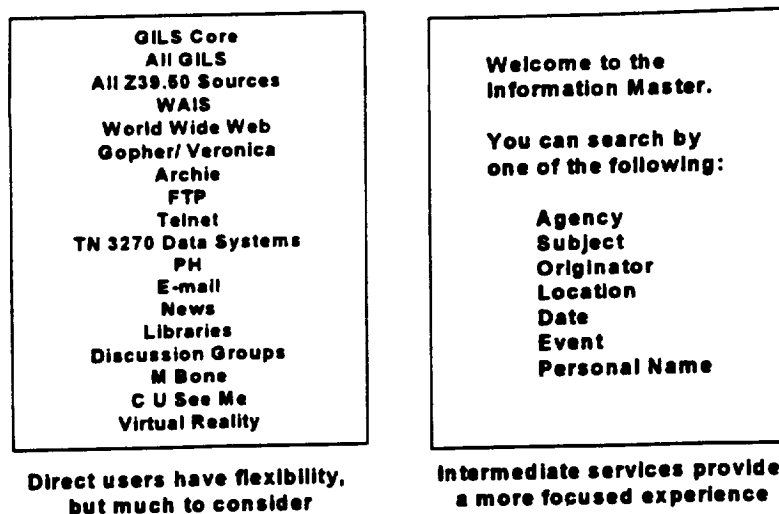


Figure 1. The public will use GILS either directly or through intermediaries.

In contrast, intermediate services are typically oriented toward a particular user community and present a more focused experience for users searching for information. Intermediate services need not require users to have sophisticated research skills or electronic network access. Government and non-government intermediaries can present GILS information in the full range of communications media and with a variety of interpretative services as appropriate for various communities. Such services can be offered via electronic mail, bulletin boards, FAX, and other media such as CD-ROM (Compact Disk-Read Only Memory), printed publications, telephone help desks, and information kiosks in public places as envisioned in the Administration's Service to the Citizen initiative.⁴

Clearly, most of the public need for access to government information will be well served through the diverse array of public and private-sector service providers. Casual users and those lacking network access will be served typically through products and services offered by agency or non-government intermediaries such as Federal depository libraries, other public libraries, and private-sector providers. These intermediaries obtain GILS information either as direct users themselves or from other intermediaries, but the extent of government information that may be provided by any particular intermediate service is not prescribed by GILS.

Having unfettered access means that the direct user takes on much more responsibility to construct a context in which the collected information is actually coherent. Accordingly, GILS has certain expectations of direct users, whether researchers or other intermediaries. Direct users of GILS must have network access, be literate in English to at least the secondary-school level, be capable of using a personal computer, and be aware of any limitations of their own hardware or software environment.

⁴Service to the Citizen Interagency Task Force. (1993). Service to the Citizen Conference Report. Washington, DC: Department of Veterans Affairs.

Data and Information

Given the huge amounts and vast range of Federal holdings, one might want to synthesize information by combining data from multiple sources as, for example, to support large scale environmental monitoring. It is important to understand that GILS operates at the level of information about data holdings. GILS addresses how to find files but does not address how the contents of those files may be accessed or used.

Users must be aware that data combined from multiple sources should be used with caution and subjected to appropriate review. Except in very strictly defined domains where common practices are rigidly enforced and data processing is well coordinated, there does not exist sufficiently detailed documentation about the data to ensure its appropriate use for purposes other than for which it was initially gathered. This situation is not peculiar to Federal holdings--whenever data is collected and maintained, it is only possible to provide for a limited set of secondary uses.

In some communities of interest, such as the participants in the National Spatial Data Infrastructure, there is strong consensus on the high secondary use value of certain basic data. This perceived value justifies large investments in data management and the establishment of multi-lateral coordination structures such as the Federal Geographic Data Committee established under OMB Circular A-16. Data management issues surrounding the international Global Change Research Program and the work of the Committee on Earth and Natural Resources are also generating some convergence of opinion on raising the level of data management investments.

While there are complex issues surrounding data comparability, it is clear that complete and readily accessible information about data holdings will be a key requirement. GILS does provide a basis for broad accessibility to the highest level description of information holdings.

The Provider Perspective

A key concept of GILS is that it uses network technology to support many different views across many separate locators.⁵ A locator is defined as an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

Although directly accessible via electronic networks such as the Internet, all or part of the GILS contents can also be made available by intermediaries through virtually any media. These alternative mechanisms help assure that the information is available through a diversity of sources, both public and private, and cover the full range of communications media from telephone help through printed publications and up to the most sophisticated electronic network technologies.

GILS organizes a collective set of agency-based locators and associated information services. Being decentralized, responsibilities can be kept close to those who understand and care for the information and who are serving the agency's primary user community. Each agency is responsible

⁵ The design of GILS follows generally a 1992 report to OMB, NARA, and the General Services Administration (GSA): McClure, Charles R., Ryan, Joe & Moen, William E. (1992). Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-based Locators 2 Vols. Bethesda, MD: National Audio Visual Center. [Available from ERIC, document no. ED349031].

for ensuring that its GILS components are continuously accessible to GILS direct users. Certain agencies, such as NARA, the Government Printing Office (GPO), and the National Technical Information Service (NTIS), also have in their primary mission an additional role in helping the public to access information maintained elsewhere in the Government. These agencies will assist in providing GILS services when requested by other agencies.

Services for finding government information take many forms, and the electronic aspects of GILS should be seen within the larger context of government information services (Figure 2). For example, the public is served through information desks in Federal buildings as well as telephone help desks and reference services such as "1-800-USA-MAPS." Many kinds of finding aids are used in such services--printed catalogs and directories are and will continue to be very common. With GILS, it will be much easier for those services to provide information drawing on the full range of Federal information resources rather than just agency-specific resources.

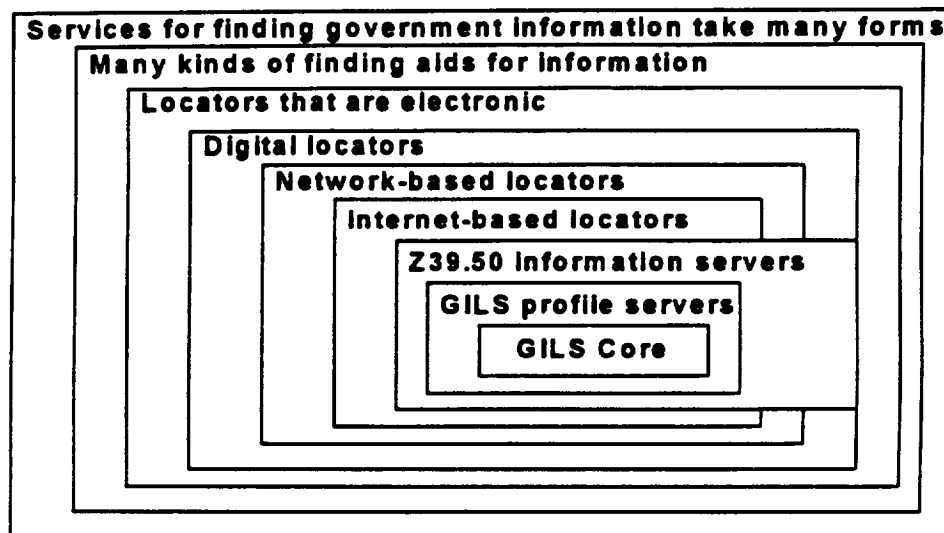


Figure 2. Electronic networks are one aspect of the Government Information Locator Service.

Among the government information finding aids are electronic media, including television announcements about government information available from the Consumer Information Center in Pueblo, Colorado. As interactive television becomes more available to homes, GILS will help to simplify the ways in which those services help the public to find Federal information resources. Also within the realm of digital electronic finding aids, there are popular information dissemination technologies such as bulletin boards and CD-ROM's. These personal, print media, and electronic services can be used to publicize GILS contents. These services may also be regarded as information resources, and may be referenced in GILS locator records themselves.

Some digital electronic finding aids use various kinds of networks and so are able to provide access to many different resources, often with a common user interface. In this area, it becomes possible to provide services in GILS where the user can have immediate access not only to information about an information resource, but to the referenced resource itself.

As stated above, GILS takes advantage of network technologies to allow many different information sources to be separately maintained yet be comprehensible as a coherent whole from the unique perspective of a specific user. This is achieved within computer networks that support peer-to-peer relationships and thereby allow for applications to operate using a client-server architecture. All of the server applications that also use the ANSI Z39.50 information search and retrieval protocol can be accessed by GILS direct users.

Because GILS uses interoperable standards for information search and retrieval, information sources referenced in GILS can be placed into virtually any context. Other major Federal government information systems such as the GPO Access System, the NTIS FedWorld system, the National Geospatial Data System, and the Global Change Data and Information System will be accessible to GILS direct users. GILS direct users may have access to a wide range of additional Federal information on the network such as current and historical information on Federal programs and institutions; public notices; law, regulation, policy, and procedural materials; and listings of experts and office locations. Agencies such as NARA, GPO, and NTIS, as well as private-sector information providers, can supplement the GILS Core with access to other Federal and non-Federal information.

Other government (state, local, tribal, foreign, international) and non-government organizations will also be encouraged to institute locators compatible with the standards used in GILS. GILS will accommodate the expressed needs of other government organizations where practical.

Design Principles

GILS is a component of the National Information Infrastructure (NII) that is evolving with guidance from the Information Infrastructure Task Force.⁶ GILS will be interoperable with other component NII initiatives such as the National Spatial Data Infrastructure. GILS is also expected to adapt to and encourage technical innovation, especially in ways that enhance public access to government information.

GILS will conform to national and international standards for information and data processing. Participants in GILS will use voluntary standards processes, e.g., ANSI, the Open Systems Environment Implementors Workshop (OIW), and the Internet Engineering Task Force, to promote interoperability of search and retrieval mechanisms, network communications, user authentication, and resource identifiers, among other essential components. Near-term implementations of GILS will use the Internet and its communications protocols, but GILS is based on the international Open Systems Interconnection (OSI) model to be compatible with a wide range of technologies. NIST, working through the OIW, will maintain and publish the application profile specifying GILS compliance.

⁶Information Infrastructure Task Force (September 15, 1993). The National Information Infrastructure: Agenda for Action. Washington, DC: NTIA NII Office, Department of Commerce. Available in ASCII text format under the file name `niiagend.asc` on the NTIA Bulletin Board (202) 482-1199 and the Fedworld bulletin board (703-321-8020). It is available on the Internet under the file name `niiagenda.asc` by anonymous FTP (File Transfer Protocol) at host `ftp.ntia.doc.gov` under the directory `/pub`, and by gopher at `gopher.nist.gov` in the menu item DOC Documents.

GILS takes advantage of the network technology known as client-server architecture, which allows locator records to be distributed among multiple independent information servers. Client applications may allow the user to question many servers concurrently and have the answers automatically combined. In this way, GILS allows for agencies to maintain GILS locator records within various information resources optimized for their usual customers, while allowing the locator information to be rapidly collated in different ways to serve different needs.

Functional Requirements

Direct users of GILS must be able to use non-proprietary standard mechanisms to discover information sources and retrieve basic textual information content. These functions are within the scope of the information search and retrieval standard known in the United States as ANSI Z39.50 and internationally as ISO (International Organization for Standardization) 10162/10163. GILS locators must be accessible on interconnected electronic network facilities and must support the currently approved ANSI Z39.50 standard for information search and retrieval. Software conforming with ANSI Z39.50 must also conform to the GILS Profile to provide full functionality to GILS direct users. In particular, the GILS Profile provides for navigating among Federal government locators through the specifications given for the GILS Core locator records. Special provisions are made in GILS to support navigation among GILS locators by using browsing as well as textual searching.

The GILS Profile provides a complete specification of GILS as it makes use of ANSI Z39.50, but also specifies where necessary those characteristics of GILS that are not within the scope of ANSI Z39.50. The GILS Profile does not limit how information is maintained at the source nor how the information is displayed to the user. Access to GILS is expected to be embedded within many different computer applications, ranging from the very simple to those that support concept searching across languages, dynamically interpret natural language, or filter search requests to sift huge amounts of information automatically. Public domain client software that supports access to GILS will be available from GPO, NTIS, and the Clearinghouse for Networked Information Discovery and Retrieval, among others.

Alternative ways to organize and present networked information are encouraged, but agencies participating in GILS will implement such alternatives in addition to supporting access by GILS direct users who employ the currently approved ANSI Z39.50 standard. For example, information organized via the OSI X.500 Directory Services standard can be made accessible also via ANSI Z39.50, thereby enhancing access capabilities. It should also be noted that a GILS direct user will typically use client software that provides access to a variety of information sources that do not comply with the GILS profile but are compliant with various other standards.

Some internal redundancy in GILS is to be expected--there will often be multiple GILS locator records describing the same resource and different search strategies applied by different intermediaries. Such redundancy is appropriate because the same information resources may be described differently to different audiences or for different purposes, and descriptions will cover information resources at a wide range of aggregation. Also, the same information resources may be described differently by different information services that participate directly or as intermediaries in providing Federal information to the public. Because GILS incorporates a

variety of automated and manual search techniques, questions will be answered from different perspectives depending on how GILS is used.

GPO (and perhaps NARA, NTIS and other agencies) will maintain a publicly accessible GILS source that provides a comprehensive directory of all GILS Core locator records from a Federal perspective. When appropriate to their respective missions, Federal agencies may also develop and maintain additional interagency topical locators that enhance opportunities for sharing information resources. The following are examples of topics that might be the subject of additional interagency locators: economic indicators, trade information, spatial data, educational and training resources, disaster relief, health information, biodiversity and global change research. Such locators would be similar in function to the GILS Core, but would not necessarily use the GILS Core Elements format nor be focused solely on Federal agency holdings.

GILS supports seamless access not only among locators but directly to referenced information resources. When implemented at both the client and the server, GILS linkages facilitate the electronic delivery of off-the-shelf information products, as well as connection to data systems that support analysis and synthesis of information (Figure 3). Although the trend is clearly in the direction of electronic network availability, much of the referenced information is not available currently in electronic form. GILS always provides information regarding request and delivery procedures for various distribution options as defined by the disseminating organization.

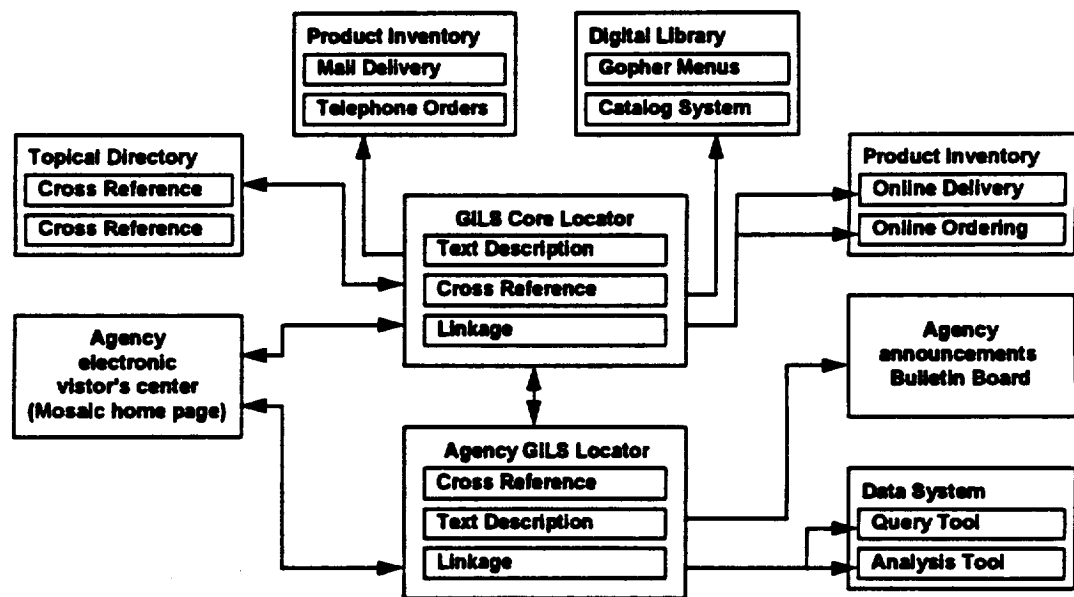


Figure 3. GILS facilitates seamless access among locators and directly to information resources.

The GILS Core

Among the GILS agency components is a set of locator records that reside on GILS accessible servers and are further identified by agencies as belonging to the GILS Core. GILS Core locator

records are required to be maintained by Federal agencies having significant information holdings, where each record describes part of the agency holdings. These Core locator records will be accessible comprehensively in the GPO Access system, but can also be aggregated by direct users of GILS to provide selective views of Federal government holdings.

The GILS Core is defined as the set of locator records maintained by the U.S. Federal government, all of which comply with the defined GILS Core Element standards, and all of which are mutually accessible through interconnected electronic network facilities. Each information disseminating agency is responsible for compiling and maintaining its own records in the GILS Core. Information services for access to GILS Core locators, once a direct user has Internet access, will be maintained by Federal agencies without charge to the direct user.

The GILS Core will include records for all information locators that catalog other publicly accessible information resources at least partially funded by the Federal government, as well as for each of the Federal government information systems that include publicly accessible data or information. While GILS Core records can point to any kind of information source, they are especially designed for helping users navigate among a wide array of other locators in various formats. It is not recommended that agencies use the precise format of the GILS Core locator records to describe all types of information resources. For example, the GILS Core Elements format would be a poor choice for describing each agency expert, but it could well be used to describe the resource that contains a compilation of such descriptions. Rather, the agency should maintain various locator records in formats appropriate to the primary user communities being served. When such other locators are published, the originating agency should include corresponding locator records that enable electronic linkage from and to the GILS Core locator.

The entire GILS Core is not likely to contain more than 100,000 locator records. In addition to locator records for information systems, it is estimated that the GILS Core will contain up to 1,000 locator records for each Federal agency that is a major disseminator of public information. Agencies that are not major disseminators will typically have fewer records in their portion of the GILS Core, especially if the agency is relatively small. Where agencies maintain information inventories that have far more records, the agency is expected to aggregate related information resources in a locator record included in the GILS Core and link the detailed inventory to GILS. Each GILS Core locator record is estimated to be less than 1,000 words in length, exclusive of any agency supplemental information that may be introduced as a separate field at the agency's discretion.

It is important to note that the vast majority of information sources accessible to GILS direct users would not be considered part of the GILS Core. Many are not maintained by the Federal Government, do not offer records in the format of the GILS Core Elements, are not on public networks, or are not offered free of charge. Many of these non-Core sources are locators nonetheless and will be very valuable to users in finding information. Also, other relevant sources of Federal information and Federal government information systems may be accessible to direct users of GILS. For example, various agencies and private-sector information providers may develop products that contain GILS Core locator records. Indeed, such derivative and value-added products may often be the first point of access to Federal information resources.

The GILS Profile

The decentralized approach envisioned for GILS requires that many different implementations be fully interoperable when implemented, although developed separately. To assure interoperability, implementors of information systems must have a clear statement of the functions of GILS and the environment within which GILS will be used. That statement becomes part of a GILS Profile that documents the specific agreements established by consensus among active implementors together with Federal representatives. The GILS Profile identifies specific standards, and the chosen subsets, options, and parameters of those standards, needed to achieve interoperability in the specific limited context of GILS.

As an initial step toward a Stable Implementors Agreement recognized by the OIW, a draft profile was created through a Cooperative Agreement between the U.S. Geological Survey and Syracuse University, with active involvement from several ANSI Z39.50 implementors representing non-government sectors. The draft GILS Profile specifies that the GILS locator records are to be available in three record syntaxes--Generic Record Syntax, United States Machine Readable Cataloging (USMARC)⁷, and Simple Unstructured Text Record Syntax (SUTRS).

When using the Generic Record Syntax, the GILS locator elements can support representation in Hypertext Markup Language (HTML). (HTML is the format interpreted by the National Center for Supercomputing Applications Mosaic client software when presenting World Wide Web objects, for example.) Provision has also been made in the GILS profile to support switching among navigation techniques, including use of a browsing mode as in Gopher or a searching mode as in bibliographic systems or Wide Area Information Servers (WAIS). The incorporation in GILS of Uniform Resource Identifiers (URIs) greatly simplifies electronic navigation among locators and other data systems available on interconnected networks.

Content definitions describe the GILS Core Elements required for users to determine the relevance of defined information resources to their needs and to understand subsequent actions to obtain the information resources (see Appendix A). These definitions identify relations among GILS Core Elements, and between GILS Core Elements and the USMARC format for bibliographic data. ANSI Z39.50 definitions of GILS Core Elements in the GILS Profile provide a structure and format for movement of the GILS Core Elements between computer systems. The Abstract Record Syntax and Basic Encoding Rules used to define GILS Core Elements are also suitable for movement of element contents between automated systems using digital media such as tape, diskette, or CD-ROM.

The GILS Profile offers a preferred display format for use in printed media as well as in electronic presentations. Although specified for human viewing in English, it is intended to be extensible to other languages also.

⁷ USMARC is an implementation of ANSI Z39.2. American National Standards Institute. (1985). American National Standard Z39.2-1985 Bibliographic Information Interchange. New York, NY: American National Standards Institute. See also USMARC Format for Bibliographic Data. Washington, DC: Cataloging Distribution Service, Library of Congress.

Appendix A. GILS Core Elements

Title: This mandatory element occurs once per locator record. It conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject. (USMARC Tag 245\$a)

Control Identifier: This mandatory element occurs once per locator record. It is defined by the information provider and is used to distinguish this locator record from all other GILS Core locator records. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual. (USMARC Tag 001)

Abstract: This mandatory element occurs once per locator record. It presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length. (USMARC Tag 520)

Purpose: This mandatory element occurs once per locator record. It describes why the information resource is offered and identifies other programs, projects, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources. (USMARC Tag 500)

Originator: This mandatory element occurs once per locator record. It identifies the information resource originator, named as in the U.S. Government Manual where applicable. (USMARC Tag 710\$a)

Access Constraints: This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any access constraints applied to assure the protection of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here. (USMARC Tag 506)

Use Constraints: This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any use constraints applied to assure the protection of privacy or intellectual property and any other special restrictions or limitations on using the information resource. (USMARC Tag 540)

Appendix A. GILS Core Elements

Availability: This mandatory element occurs one or more times per locator record. It is a grouping of sub-elements that together describe how the information resource is made available.

Distributor: This mandatory sub-element occurs once per Availability element. It identifies the distributor by name, organization, street address, city, state, zip code, country, network address, hours of service, telephone, and/or fax number. (USMARC Tag 037\$b)

Resource Description: This optional sub-element occurs no more than once per Availability element. It identifies the resource as it is known to the distributor. (USMARC Tag 037\$f)

Order Process: This mandatory sub-element occurs once per Availability element. It provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods. (USMARC Tag 037\$c)

Technical Prerequisites: This optional sub-element occurs no more than once per Availability element. It describes any technical prerequisites for use of the information resource as made available by this distributor. (USMARC Tag 538)

Available Time Period: This optional sub-element may occur multiple times per Availability element. It provides the time period reference for the information resource as made available by this distributor. (Time period formats are as given for the Time Period of Content element described below.)

Available Linkage: This optional sub-element occurs no more than once per Availability element. It provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information. (USMARC Tag 856\$u)

Available Linkage Type: This optional sub-element occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI. (USMARC Tag 856 first indicator/ 856\$2)

Point of Contact for further information: This mandatory element occurs once per locator record. It identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact. Defined sub-elements include name, organization, street address, city, state, zip code, country, network address, hours of service, telephone, and fax number. (USMARC Tag 856\$m for electronic resources, 535 for non-electronic resources)

Record Source: This mandatory element occurs once per locator record. It identifies the organization, as named in the U.S. Government Manual, that created or last modified this locator record. (USMARC Tag 040)

Appendix A. GILS Core Elements

Date Last Modified: This mandatory element occurs once per locator record. It identifies the latest date on which this locator record was created or modified. (USMARC Tag 005)

Agency Program: This element occurs no more than once per locator record. It identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource. This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system. (USMARC Tag 500)

Sources of Data: This element occurs no more than once per locator record. It identifies the primary sources or providers of data to the system, whether within or outside the agency. This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system. (USMARC Tag 500)

Controlled Vocabulary: This optional element may occur multiple times per locator record. It is a grouping of sub-elements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary.

Index Terms - Controlled: This sub-element occurs once per Controlled Vocabulary element. It is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field, Controlled Term. (USMARC Tag 650)

Thesaurus: This sub-element occurs once per Controlled Vocabulary element. It provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. (USMARC Tag 650 first indicator/ 650\$2)
Notes on how to obtain electronic access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.

Local Subject Index: This optional element occurs no more than once per locator record. It is a grouping of descriptive terms to aid users in locating resources of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating sub-element, Local Subject Term. (USMARC Tag 653\$a)

Methodology: This optional element occurs no more than once per locator record. It identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described. (USMARC Tag 567)

Appendix A. GILS Core Elements

Spatial Reference: This optional element occurs no more than once per locator record and provides the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

Bounding Rectangle: This optional sub-element occurs no more than once within a Spatial Reference element. It provides the limits of coverage expressed by latitude and longitude values in the order: western-most, eastern-most, northern-most, southern-most.

(USMARC Tags 255\$c, 034\$d, 034\$e, 034\$f, 034\$g)

Geographic Name: This optional sub-element may occur multiple times within a Spatial Reference element. It identifies significant areas and/or places within the coverage through two associated constructs: a Geographic Keyword Name (USMARC Tag 651) and a Geographic Keyword Type (USMARC Tag 655). A preferred source of the names and types is the Geographic Names Information System.

Time Period of Content: This optional element may occur multiple times per locator record. It provides time frames associated with the information resource, in one of two forms:

Time period - structured: Time described using the USMARC prescribed structure. (USMARC Tag 045\$c)

Time period - textual: Time described textually. (USMARC Tag 513)

Cross Reference: This optional element may occur multiple times per locator record. Each instance is a grouping of sub-elements that together identify another locator record likely to be of interest.

Cross Reference Title: This optional sub-element occurs no more than once per Cross Reference element. It provides a human readable textual description of the cross reference. (USMARC Tag 787\$t)

Cross Reference Linkage: This optional sub-element occurs no more than once per Cross Reference element. It provides the machine readable information needed to perform the access (i.e., URI). (USMARC Tag 787\$w)

Cross Reference Type: This optional sub-element occurs if there is a Cross Reference Linkage described. It provides the data content type (i.e., MIME) for the referenced URI. (USMARC Tag 856 first indicator/ 856\$2)

Original Control Identifier: This optional element occurs no more than once per locator record. It is used by the record source to refer to another GILS locator record from which this locator record was derived. (USMARC Tag 035)

Supplemental Information: This optional element occurs no more than once per locator record. Through this element, the record source may associate other descriptive information with the GILS Core locator record. (USMARC Tag 500)

Appendix A. GILS Core Elements

Appendix B: Glossary

agency - any executive department, military department, government corporation, government controlled corporation, or other establishment in the executive branch of the United States Federal government, or any independent regulatory agency (OMB Circular A-130).

ANSI Z39.50 - The "American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection" is developed by the National Information Standards Organization (NISO), accredited to the American National Standards Institute (ANSI). ANSI Z39.50 complies with the Open Systems Interconnection (OSI) family of standards promulgated by the International Organization for Standardization (ISO), and is interoperable with the international standards for information search and retrieval, ISO 10162 and 10163. As of this writing, the currently approved version is ANSI Z39.50 Version 2.

direct user - a person or automated process that accesses GILS from networks using the GILS Profile and thereby having more flexibility to explore the full complement of available information. People who are direct users of GILS are assumed to be literate in English to at least the secondary school level, capable of using a personal computer, and aware of any constraints of their own hardware or software environment.

dissemination - the government initiated distribution of information to the public, excluding distribution limited to government employees or agency contractors or grantees, intra-agency or inter-agency use or sharing of government information, and responses to requests for agency records under the Freedom of Information Act (5 U.S.C. 552) or Privacy Act. Here, "disseminating information" is not distinguished from "providing access to information" (following OMB Circular A-130).

electronic information resource - information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

government information - information created, collected, processed, disseminated, or disposed of by or for the Federal government (OMB Circular A-130).

Government Information Locator Service (GILS) - a decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

GILS Core - a subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government, comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

government publication - information that is published as an individual document at government expense, or as required by law (OMB Circular A-130).

Appendix B. Glossary

information - any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms (OMB Circular A-130).

information product- any book, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic (OMB Circular A-130).

information resource - includes both government information and information technology (OMB Circular A-130).

information service - considered equivalent to information product from the policy perspective of OMB Circular A-130, although agency locator records for services may differ from those for products.

information system - the organized collection, processing, maintenance, transmission, and dissemination of information in accordance with defined procedures, whether automated or manual (OMB Circular A-130).

information technology - the hardware and software operated by a Federal agency or by a contractor of a Federal agency or other organization that processes information on behalf of the Federal Government to accomplish a Federal function (OMB Circular A-130).

intermediary or intermediate service - an entity or service that makes some of the GILS information available but does not provide the full capabilities of a direct user.

interoperability - a condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

locator - an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

locator record - a collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information.

mandatory element - a data element in a GILS Core Locator Record that must have a value provided by the record source.

Open Systems Interconnection (OSI) - a family of standards promulgated by the International Organization for Standardization (ISO) and adhering to a specific model that promotes interoperability.

profile - the statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters of those standards; a set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

Appendix B. Glossary

records management - the planning, controlling, directing, organizing, training, promoting, and other managerial activities involved with respect to records creation, records maintenance and use, and records disposition in order to achieve adequate and proper documentation of the policies and transactions of the Federal government and effective and economical management of agency operations. (44 U.S.C. 2901(2))

Uniform Resource Identifier (URI) - a set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC - an implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

Appendix B. Glossary

APPLICATION PROFILE FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE(GILS)

1. Introduction

This document describes an application profile for the Government Information Locator Service (GILS). The GILS Profile includes not only the specifications for ANSI/NISO Z39.50, the American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (National Information Standards Organization, 1992) in the application but also other aspects of a GILS conformant server that are outside the scope of Z39.50. The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core, which is a subset of all GILS Locator Records, and completely specifies the use of Z39.50 in this application.

2. Background

The GILS is a response to the need for users to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. Christian (1994) is the authoritative document providing an overview of GILS, its objectives, service requirements, and core requirements. According to Christian (1994), the GILS is an overall service and includes information and technology components as well as policy, regulation, people, etc. The GILS is intended to help the public locate and access public information throughout the U.S. government.

The current GILS initiative builds upon a previous study, Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992). That study, which was conducted for the Office of Management and Budget, the National Archives and Records Administration, and the General Services Administration, recommended that each agency establish a network-accessible locator that describes its information resources. The study also recommended that agencies use Z39.50 as the appropriate information retrieval protocol to achieve a distributed, standards-based Government Information Locator Service.

The development of the GILS Profile is documented in Using Z39.50 in an Application for the Government Information Locator Service (GILS) (McClure & Moen, 1994). The GILS Profile resulted from the work of a group comprising experts in Z39.50 implementations, system implementations, and information organization, and representatives of Federal agencies. The specifications included in the GILS Profile reflect the consensus of this group and input from a range of stakeholders.

3. Scope

The GILS Profile fully specifies the use of ANSI/NISO Z39.50 by the GILS. In addition, the GILS Profile provides the specifications for the overall GILS application relating to the GILS Core including other aspects of GILS conformant servers that are outside the scope of Z39.50.

This version of the GILS Profile focuses on requirements for a GILS server operating in the Internet environment. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner that allows interoperability with the GILS server. Clients that support Z39.50 but do not implement the GILS Profile will be able to access GILS records with less than full GILS functionality.

The GILS Profile addresses many aspects of the GILS (e.g., intersystem interactions and information interchange) but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality.

4. Field of Application

The GILS Profile supports search and retrieval of GILS Locator Records contained in GILS servers by users in the Internet environment.

The GILS Profile will be used by developers of GILS servers. It will also be used by client developers to understand expected behaviors of GILS servers. A GILS server accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. Some of these information resources pointed to by GILS Locator Records, as well as the GILS server itself, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail (SMTP/MIME). The use of these protocols or other communications paths is outside the scope of the GILS Profile.

Once connected to a GILS server, users supported by appropriate clients that understand the GILS Profile may navigate through single or multiple servers. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of these activities (e.g., browsing is also a client function in the sense of how it interprets and presents GILS data).

5. References

The following list contains documents that contain provisions which, through reference in this text, constitute provisions of the GILS Profile. At the time of this publication, the editions indicated were valid. All documents are subject to revision, and parties to agreements based on this Profile are warned against automatically applying any more recent editions of the documents listed below, since the nature of references made by the Profile to such documents, is that they may be specific to a particular edition. In addition, this list contains other documents that can be consulted for further information, background, etc.

- [1] American National Standards Institute. (1985). American National Standard 39.2-1985 Bibliographic Information Interchange. New York: American National Standards Institute.
- [2] Christian, Eliot. (1994, May 2). Government Information Locator Service (GILS): Report Information Infrastructure Task Force. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).
- [3] Lynch, Clifford A. (1994, April 30). "Using the Z39.50 Information Retrieval Protocol in the Internet Environment" [Draft RFC for Z39.50 over TCP/IP].
- [4] McClure, Charles R. & Moen, William E. (1994, May 7). Using Z39.50 in an Application for the Government Information Locator Service (GILS). Available via anonymous FTP at <ericir.syr.edu> as /USGS/profile_background.doc.ps (Postscript format) and as /USGS/profile_background.doc.txt (ASCII text format).
- [5] McClure, Charles R., Ryan, Joe & Moen, William E. Moen. (1992). Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators. 2 Vols. Bethesda, MD: National Audio Visual Center [Available from ERIC, document no. ED349031].
- [6] National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Gaithersburg, MD: NISO Press.

- [7] National Institute of Standards and Technology. (1992). FIPS No. 173, Spatial Data Transfer Standard (August 28, 1992). Gaithersburg, MD: National Institute of Standards and Technology.
- [8] Office of Management and Budget. (1993). Circular No. A-130, "Management of Federal Information Resources" (58 F.R. 36068, July 2, 1993).
- [9] Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA). (1993). OIW/SIGLA Document #1: Using Z39.50-1992 Directly over TCP.
- [10] RFC 1521, MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies.
- [11] RFC 1522, MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text.
- [12] Uniform Resource Locators (URL): A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network. (1993, October). [Internet Draft]. The latest URL draft is:
<url:ftp://info.cern.ch/pub/www/doc/url7a.txt>
- [13] Uniform Resource Names. (1993, October). [Internet Draft]. The latest URN draft is:
<url:ftp://ds.internic.net/internet-drafts/draft-ietf-uri-resource-names-01.txt>
- [14] USMARC Format for Bibliographic Data. Washington, DC: Library of Congress, Cataloging Distribution Service.

6. Definitions

For purposes of this Profile, the following definitions apply.

Client: An initiating application. This application includes the Z39.50 origin.

Electronic Information Resource: Information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

GILS Core: A subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government and comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

Government Information: Information created, collected, processed, disseminated, or disposed of by or for the Federal government.

Government Information Locator Service (GILS) : A decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

Information Resource: Includes both government information and information technology.

Interoperability: A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

Locator Record: A collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information.

Mandatory: An element in a GILS Core Locator Record that must have a value provided by the record source. The GILS Profile does not specify which elements must be present from the perspective of GILS servers.

Origin: The part of a client application that initiates a Z39.50 association and is the source of requests during the association.

Profile: The statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters of those standards. A set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

Registered Object: An object that is identified by a name-to-thing relationship in which the name is recorded by a registration authority to ensure that the names can be used unambiguously.

Server: An application that responds to an initiating application (i.e., a client). The application that includes the Z39.50 target.

Target: The part of an server application that accepts a Z39.50 association.

Uniform Resource Identifier (URI): A set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC: An implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

7. Z39.50 Specifications for GILS

This section details the required services available from Z39.50, describes an Attribute Set for searching, four Element Set Names by which the server presents some or all the elements (defined in the Schema) of the Locator Records, and prescribes the Record Syntaxes to be supported by GILS servers for the transfer of Locator Records.

7.1. Version

GILS clients and servers support Z39.50 Version 2 as specified in Z39.50-1994. GILS requires support of various objects, some of which are not defined in Z39.50-1992. These are listed in 7.2.

7.2. GILS Objects

The following object identifier (OID) is assigned to the Z39.50 standard:

{iso (1) member-body (2) US (840) ANSI-standard-Z39.50 (10003)}.

This OID is abbreviated as: ANSI-standard-Z39.50.

Several object classes are assigned at the level immediately subordinate to ANSI-standard-Z39.50, including:

- 3 = attribute set definitions
- 4 = diagnostic definitions
- 5 = record syntax definitions
- 13 = database schema definitions.
- 14 = tagSet definitions.

GILS requires support of the following objects

- | | |
|-------------------------|------------------------------|
| • GILS attribute set: | {ANSI-standard-Z39.50 3 3} |
| • bib1 diagnostic set: | {ANSI-standard-Z39.50 4 1} |
| • USMARC record syntax: | {ANSI-standard-Z39.50 5 10} |
| • SUTRS record syntax: | {ANSI-standard-Z39.50 5 101} |
| • GRS-1 record syntax: | {ANSI-standard-Z39.50 5 105} |
| • GILS schema: | {ANSI-standard-Z39.50 13 2} |
| • tagSet-M | {ANSI-standard-Z39.50 14 1} |
| • tagSet-G | {ANSI-standard-Z39.50 14 2} |

7.3. Communication Services

When Transmission Control Protocol (TCP) is used as the transport service, the specification for use of TCP is found in OIW/SIGLA Document #1, "Using Z39.50-1992 Directly over TCP." The use of other communication services is not yet defined.

7.4. Z39.50 Services

There are three Z39.50 (Version 2) services that are required for conformance: Init, Search, and Present. No additional services are required for conformance to the GILS Profile. Other Z39.50 services, however, may be provided optionally by servers and used by clients.

Standard Z39.50 Init Service negotiation procedures control the use of all services.

7.4.1. Search

The GILS application will support Z39.50 Type 1 queries which are general purpose Boolean query structures.

7.4.1.1. Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute set and consists of all Bib-1 Attributes and additional Use Attributes that are defined for GILS elements (see Annex A for the GILS Use Attributes). These newly defined GILS Use Attributes are well-known and correspond semantically to GILS Core Elements. The GILS Attribute Set is a registered object.

GILS servers must support a limited number of GILS Attributes. The required GILS Attributes follow. (Note: The GILS Use Attribute is listed followed by the GILS Use Attribute Number and the corresponding GILS Core Element):

- o Use Attributes: Local Number (12; Local Control Number); Author-name corporate (1005; Originator); Date/Time Last Modified (1012; Date of Last Modification); Record Source (1019; Record Source); Distributor Name (2001; Distributor Name); Index Terms -- Controlled (2002; Index Terms -- Controlled); Local Subject Index (29; Local Subject Term); Any (1016)
- o Structure: Word (2), URx (104), Date (5), Word List (6)
- o Relation: Greater than (5), Equal (3).

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of required GILS Attributes listed in Table 1 in Annex A.

7.4.1.2. Well-known Search

To provide support for browsing GILS Locator Records, there is a well-known search consisting of the following GILS Attributes: Use Attribute: Local Number; Structure Attribute: URX; and a term of zero length. GILS servers that support browsing of records will create a result set of one or more GILS Locator Records that provide the necessary information to allow clients to offer menu-like displays of GILS Locator Records or other information and information resources.

The "Browse" in the GILS context involves only the Search and Present Services of Z39.50. "Browse" is used informally in the GILS Profile, and it is not related nor should it be confused with the Browse Facility or Scan Service of Z39.50.

7.4.2. Retrieval

This section describes the components and procedures used by Z39.50 to return records in response to a query.

7.4.2.1. Schema

The GILS Profile specifies a GILS Schema (see Annex D for the Schema). The GILS Schema is a registered object. The schema describes and/or defines tagSets used and an abstract record structure for a Locator Record. A schema in Z39.50 can be modified and may evolve over time, and it is reasonable to expect the GILS Schema will evolve.

The GILS Schema uses elements from tagSet-M and tagSet-G and defines in the GILS tagSet additional elements as necessary. The GILS Profile specifies tagTypes to identify tagSet-M elements (tagType = 1), tagSet-G elements (tagType = 2), and the elements defined by the GILS tagSet (tagType = 4). Another tagType (tagType=3) is used to identify arbitrary string tags for locally defined elements.

The GILS tagSet element numbering begins with number 1. Elements can be nested and the tagging notation (i.e., the tag path) will reflect the nesting.

All well-known GILS Schema elements have assigned numeric tags. String-tags (i.e., text) may be used in the GILS Schema to label those elements that are not well-known (i.e., locally defined).

7.4.2.2. Element Sets Names

GILS servers will support four Element Set Names. GILS servers will interpret the use of the Element Set Names required by the GILS Profile to identify the following elements from the GILS Schema:

- The primitive element set name "B" contains at least: title, controlIdentifier, originator, and local control number
- The primitive element set name "G" contains: all B Element Set elements and crossReference
- The primitive element set name "W" contains: all B Element Set elements and bodyOfDisplay.

- The primitive element set name "F" contains: all elements available in the record.

The server should include in a retrieved record all of the elements specified by the element set name for which there is data available in the database record and which can be encoded in the requested record syntax (e.g., some types of locally defined binary data may not be encodable in a USMARC or SUTRS record).

7.4.2.3. Record Syntaxes

GILS servers are required to support the following three record syntaxes:

- USMARC -- an implementation of ANSI/NISO Z39.2 and maintained by the Library of Congress
- Generic Record Syntax (GRS-1) -- defined in Z39.50
- Simple Unstructured Text Record Syntax (SUTRS) -- defined in Z39.50.

Annex B contains a mapping of Core Elements to USMARC for use in the USMARC record syntax. However, since the data transformation is not fully reversible and requires interpretation, the record source is responsible for encoding the USMARC record(s).

The data in GILS Locator Records do not always map clearly into USMARC records, particularly when agencies add their own locally defined fields to the GILS Locator Record. This means that construction of USMARC records is subject to local interpretation. Therefore, GILS Locator Records in USMARC format obtained from other than the original record source should be considered non-definitive. The original source of the GILS Locator Record can be identified by examining the Original Control Identifier field of the record.

For interchange, GRS-1 records are to be treated as the complete and canonical representation; SUTRS and USMARC should be viewed as derivative records from the canonical representation and as such are not as complete or precise.

7.5. Preferred Display Format for Use with SUTRS

The GILS Profile recommends a preferred display format for SUTRS records (see Annex C for the recommended display format). For the SUTRS records, formatting instructions for a preferred display format is a concern of the server.

When the target transfers a GILS record using the SUTRS record syntax, it will encode the GILS record formatted according to the preferred display format, so that the client may present the record directly, without processing. For SUTRS, however, the client should not expect to be able to parse the record to obtain any individual GILS elements.

When the client presents a GILS record formatted by the server using the USMARC or GRS record syntax, it is recommended that the client consider the SUTRS suggested display layout in formatting the received record for presentation to the human end user.

7.6. Diagnostic Messages

The GILS application will use Diagnostic Set Bib-1.

8. Data Elements in the Locator Records

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The GILS Core Elements are defined in Annex E.

Annex A

GILS Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute Set and consists of all Bib-1 Attributes and the additional Use Attributes listed below. Additional Use Attributes that cannot be mapped to Bib-1 Use Attributes are numbered from 2000 through 2999. These are well-known Use Attributes.

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of GILS Attributes listed in Table 1. An "X" in the table means that GILS servers will recognize and support this combination of Attributes.

USE	WORD	URx	DATE	W O R D LIST	GREATER THAN	EQUAL
Local Number	X	X		X		X
Author-name corporate	X			X		X
Date /Time Last Modified			X		X	X
Record Source	X			X		X
Distributor Name	X			X		X
Index Term -- Controlled	X			X		X
Local Subject Index	X			X		X
Any	X			X		X

TABLE 1
Recognized and Supported Combinations of GILS Attributes

As stated in 7.3.1.1, GILS servers are required to support a minimal set of Use Attributes. These are listed first. In the cases where a Bib-1 Use Attribute's Name is used, the corresponding GILS Core Element name appears in parentheses.

Required GILS Use Attributes

<u>Use #</u>	<u>GILS Attribute Name</u>
12	Local Number (Local Control Number)
29	Local Subject Index (Local Subject Term)
1005	Author-name corporate (Originator)
1012	Date/Time Last Modified (Date of Last Modification)
1016	Any
1019	Record Source
2001	Distributor Name
2002	Index Terms -- Controlled

Available GILS Use Attributes

<u>Use #</u>	<u>GILS Attribute Name</u>
4	Title
1007	Identifier - Standard (Control Identifier)
62	Abstract
2003	Purpose
2004	Access Constraints
2005	Use Constraints
2006	Distributor Organization
2007	Distributor Street Address
2008	Distributor City
2008	Distributor State
2010	Distributor Zip Code
2011	Distributor Country
2012	Distributor Network Address
2013	Distributor Hours of Service
2014	Distributor Telephone
2015	Distributor Fax
2016	Available Resource Description
2017	Available Order Process
2018	Available Technical Prerequisites
2019	Available Time Period -- Structured
2020	Available Time Period -- Textual
2021	Available Linkage
2022	Available Linkage Type
2023	Contact Name
2024	Contact Organization

Available GILS Use Attributes

<u>Use #</u>	<u>GILS Attribute Name</u>
2025	Contact Street Address
2026	Contact City
2027	Contact State
2028	Contact Zip Code
2029	Contact Country
2030	Contact Network Address
2031	Contact Hours of Service
2032	Contact Telephone
2033	Contact Fax
2034	Agency Program
2035	Sources of Data
2036	Thesaurus
2037	Methodology
2038	Bounding Rectangle -- Western-most
2039	Bounding Rectangle -- Eastern-most
2040	Bounding Rectangle -- Northern-most
2041	Bounding Rectangle -- Southern-most
2042	Geographic Keyword Name
2043	Geographic Keyword Type
2044	Time Period - Structured
2045	Time Period - Textual
2046	Cross Reference Title
2047	Cross Reference Linkage
2048	Cross Reference Type
2049	Original Control Identifier
2050	Supplemental Information

Annex B

GILS Core Element to USMARC Mapping

This Annex provides a mapping from GILS Core Elements to USMARC for use by the record source and GILS servers. Some of these data elements consist of two or more subelements, and this relationship is noted by the indentation.

Implementors should consult the authoritative documentation on USMARC found in USMARC Format for Bibliographic Data. The document is available from the Cataloging Distribution Service at the Library of Congress. A full description of the USMARC fields and available subfields within each field is in that document.

For some elements new USMARC fields and/or subfields may be incorporated into the USMARC format. New fields and/or subfields in the process of being considered for inclusion in USMARC are noted.

In cases where the 500 Note field is repeated to carry separate GILS Core Elements, the name of the GILS Core Element will be included and precede the data content for that field. A colon will separate the GILS Data Element name from the rest of the content in the field. For example, 500 Purpose: [data for this field]; 500 Agency Program: [data for this field]. Each such GILS Core Element should be carried in separate, repeating 500 fields.

In addition to the variable length fields listed in the mapping, a USMARC record will also include a Leader and field 008: Fixed-Length Data Elements. Certain character positions in each of these fixed length fields of a USMARC record will need to be coded specifically for GILS. In addition, USMARC records for GILS will include a code in the 042: Authentication Code to identify these USMARC records specifically as GILS Locator Records. The following suggest values for these fields (or parts of these fields):

Leader: A fixed field comprising the first 24 character positions (00-23) of each record that provides information for the processing of the record. For GILS records, the following character position is specifically relevant:

Character Position: 18 -- Descriptive cataloging form

Value: # [i.e., blank] (Non-ISBD)

to indicate when International Standard Bibliographic Description is not followed.

008 Fixed Length Data Elements: Forty character positions (00-39) containing positionally-defined data elements that provide coded information about the record as a whole or about special bibliographic aspects of the item being cataloged. For GILS records that describe electronic information resources, the following character position is specifically relevant:

Character Position: 26 -- Type of computer file

Values: a (Numeric data)
b (Computer program)
c (Representational)
d (Document)
e (Bibliographic data)
f (Font)
g (Game)
h (Sound)
i (Online system or service) [new code proposed]
m (Combination)
u (Unknown)
z (Other)

042 Authentication Code

Value: gils [new code proposed]

GILS Data Elements and Corresponding USMARC Tags

<u>GILS Data Element</u>	<u>USMARC Tag</u>
Title	245\$a
Control Identifier	001
Abstract	520
Purpose	500
Originator	710\$a
Access Constraints	506
Use Constraints	540
Distributor	
Distributor Name	270\$p [proposed field]
Distributor Organization	270\$p [proposed field]
Distributor Street Address	270\$a [proposed field]
Distributor City	270\$b [proposed field]
Distributor State	270\$c [proposed field]
Distributor Zip Code	270\$e [proposed field]
Distributor Country	270\$d [proposed field]
Distributor Network Address	270\$m [proposed field]
Distributor Hours of Service	301\$a [proposed field]
Distributor Telephone	270\$k [proposed field]
Distributor Fax	270\$l [proposed field]
Available Resource Description	037\$f
Available Order Process	037\$c
Available Technical Prerequisites	538
Available Time Period -- Structured	045\$c
Available Time Period -- Textual	037\$n [proposed field] (for non-electronic resource)
	856\$z (for electronic resource)
Available Linkage	856\$u
Available Linkage Type	856 1st indicator/856\$2
Point of Contact	856\$m (for electronic resources)
Contact Name	270\$p [proposed field]
Contact Organization	270\$p [proposed field]
Contact Street Address	270\$a [proposed field]
Contact City	270\$b [proposed field]
Contact State	270\$c [proposed field]
Contact Zip Code	270\$e [proposed field]
Contact Country	270\$d [proposed field]
Contact Network Address	270\$m [proposed field]
Contact Hours of Service	301\$a [proposed field]

GILS Data Elements and Corresponding USMARC Tags

<u>GILS Data Element</u>	<u>USMARC Tag</u>
Contact Telephone	270\$k [proposed field]
Contact Fax	270\$l [proposed field]
Record Source	040
Date Last modified	005
Agency Program	500
Sources of Data	537 [proposed field]
Index Terms – Controlled	650
Thesaurus	650 1st indicator/ 650\$2
Local Subject Term	653\$a
Methodology	567
Spatial Reference	
Bounding Rectangle	255\$c
Western-most	034\$d
Eastern-most	034\$e
Northern-most	034\$f
Southern-most	034\$g
Geographic Name	
Geographic Keyword Name	651
Geographic Keyword Type	655
Time Period – Structured	045\$c
Time Period – Textual	513
Cross Reference Title	787\$t
Cross Reference Linkage	787\$w
Cross Reference Type	856 1st indicator/856\$2
Original control identifier	035
Supplemental information	500

USMARC Tags and Field Names**(from USMARC Format for Bibliographic Data)**

<u>USMARC Tag</u>	<u>Subfield</u>	<u>Field Name</u>
001		Control Number
005		Date and Time of Latest Transaction
034		Coded Cartographic Mathematical Data
	\$d	Coordinates - westernmost longitude
	\$e	Coordinates -- easternmost longitude
	\$f	Coordinates -- northernmost latitude
	\$g	Coordinates -- southernmost latitude
035		System Control Number
037		Source of Acquisition
	\$b	Source of stock number/acquisition
	\$c	Terms of availability
	\$f	Form of issue
	\$n	Note [proposed]
040		Cataloging Source
042		Authentication Code
245		Title Statement
	\$a	Title
255		Cartographic Mathematical Data
	\$c	Statement of coordinates
270	\$a	Address
270	\$b	City
270	\$c	State or province
270	\$d	Country
270	\$e	Postal code
270	\$k	Telephone number
270	\$l	Fax number
270	\$m	Electronic mail address
270	\$p	Contact person
301	\$a	Hours
500		General Note
506		Restrictions on Access Note
513		Type of Report and Period Covered Note
520		Summary, Etc. Note
537		Source of Data Note [proposed]
538		System Details Note
540		Terms Governing Use and Reproduction Note
567		Methodology Note

USMARC Tags and Field Names**(from USMARC Format for Bibliographic Data)**

<u>USMARC Tag</u>	<u>Subfield</u>	<u>Field Name</u>
650		Subject Added Entry -- Topical Term
1st indicator		Level of subject
	\$2	Source of heading or term
651		Subject Added Entry -- Geographic Name
653		Index Term -- Uncontrolled
	\$a	Uncontrolled term
655		Index Term -- Genre/Form
710		Added Entry -- Corporate Name
	\$a	Corporate name or jurisdiction name as entry element
787		Nonspecific Relationship Entry
	\$t	Title
	\$w	Record Control Number
856		Electronic Location and Access
1st indicator		Access method
	\$m	Contact for access assistance
	\$u	Uniform Resource Locator
	\$z	Nonpublic note
	\$2	Source of access

Annex C

Preferred Display Format for GILS Records

GILS servers will transfer records in three record syntaxes:

- o USMARC
- o Generic Record Syntax (GRS)
- o Simple Unstructured Text Record Syntax (SUTRS).

In SUTRS, the formatting of the record contents is handled by the server, and the client receives a record devoid of structure. In USMARC and GRS, the record, whose structure is defined by the record syntax, is passed from the target to an origin, and the client software has more flexibility in processing the record contents for display.

The recommended guidelines in this Annex describe how records should be displayed, whether formatted by the server or the client (but this does not preclude display formats in addition to the Preferred Display Format).

Record Organization:

The record should be organized so that the elements first viewed by the user provide adequate information to either choose or eliminate the record from further consideration. These elements are: Title, Originator, Controlled Vocabulary, Local Subject Index and Abstract.

Next in the order of presentation are elements that give detailed information about the information resource being described: Spatial Reference, Time Period, Availability, Sources of Data, Methodology, Access Constraints, Use Constraints, Point of Contact, and Supplemental Information.

The elements describing the reason for the existence of the data are next: Purpose and Agency Program.

Related information resources are listed next in the element: Cross Reference.

The final elements provide bibliographic control information: Control Identifier, Record Source, and Date of Last Modification.

General Instructions for Formatting Full Element Set Name Records:

All displayable elements are to be labelled with the full title of the field followed by a colon. Label mnemonics should only be used in situations where the user can ask for an explanation of the mnemonic. Mnemonics should not be used in SUTRS records, since it should be assumed that the client knows nothing about the server and is incapable of interpreting the mnemonics.

The subelements of constructed elements (i.e., locally defined fields, Availability, Spatial Reference, etc.) should be indented to reflect their association and structure within a well-structured element. Labels on subelements can eliminate the redundant leading parts (e.g., the word Available on the Availability subelements).

In the Controlled Vocabulary element, the Thesaurus subelement can be presented in parentheses, followed by the Index Terms. Multiple Index Terms should be separated by a semi-colon and a space (e.g., Controlled Vocabulary (MeSH): Kidney; Kidney Disease). Alternatively, the Thesaurus and Index Terms can be indented under the Controlled Vocabulary label, as is done with the other well-structured fields. Local Subject Terms should be separated by a semi-colon and a space.

Display Format for Brief Element Set Name Records:

Brief Records consist of the Title, Control Identifier, Originator, and Local Control Number fields. For display purposes, the Control Identifier and Local Control Number can be omitted. Brief Records may be formatted to fit on a single line. This may require that that one or both of the displayed fields will be truncated. Truncation can be indicated with with elipsis(...).

Display Format for G Element Set Name Records:

G Records consist of Brief Record elements and additionally, the Cross Reference element. For display purposes, the guidelines for Full Records should be followed.

Annex D

GILS Schema

The GILS Schema describes and defines tagSets and an Abstract Record Structure used with the Generic Record Syntax (GRS). The GILS Schema defines a GILS tagSet that associates a numeric tag with one or more GILS Core Elements.

Some GILS Core elements correspond to tags already defined in tagSet-M and tagSet-G, and these tags are used to identify GILS Core elements in the Abstract Record Structure. When the tagType is 1, the tag value is from tagSet-M. When the tagType is 2, the tag value is from tagSet-G. When the tagType is 3, the tag value is an arbitrary string tag. When the tagType is 4, the tag value is from the GILS tagSet.

There are two general classes of schema elements in the GILS Schema:

- 1) Primitive -- these elements cannot have locally defined subelements
- 2) Constructed -- these elements have one or more subelements any of which may be well-defined or target-defined; in the latter case, these locally defined subelements are identified with string tags

This Annex first presents the GILS tagSet that identifies the element, its unique tag, and a recommended datatype. This is followed by the GILS Abstract Record Structure that shows the full tag path for each element.

GILS tagSet

<u>Tag</u>	<u>Element</u>	<u>Recommended Datatype</u>
1	controlIdentifier	InternationalString
2	streetAddress	InternationalString
3	city	InternationalString
4	state	InternationalString
5	zipcode	InternationalString
6	hoursOfService	InternationalString
7	resourceDescription	InternationalString
8	technicalPrerequisites	InternationalString
9	westernMost	intUnit
10	easternMost	intUnit
11	northernMost	intUnit
12	southernMost	intUnit
13	geographicKeywordName	InternationalString
14	geographicKeywordType	InternationalString
15	timePeriodStructured	GeneralizedTime
16	timePeriodTextual	InternationalString
17	linkage	InternationalString
18	linkageType	InternationalString
19	recordSource	InternationalString
20	controlledTerm	InternationalString
21	thesaurus	InternationalString
22	localSubjectTerm	InternationalString
23	originalControlIdentifier	InternationalString

NOTE: The element "wellKnown" from tagSet-M (1,19) and referred to below has the following definition:

When an element is defined to be "structured into locally defined elements", the target may use this tag (i.e., wellKnown) in lieu of, or along with, locally defined tags. For example, an element named 'title' might be described to be "locally structured." The target might present the element structured into the following subelements: 'wellKnown', 'spineTitle', and 'variantTitle', where the latter two tags are target defined. In this case, 'wellKnown' is assumed to mean 'title'.

- 50 title Constructed as follows --
 This element may include the element wellKnown and may also include locally defined elements.
- 51 purpose Constructed as follows --
 This element may include the element wellKnown and may also include locally defined elements.

GILS tagSet

<u>Tag</u>	<u>Element</u>	<u>Recommended Datatype</u>
52	originator This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
53	accessConstraints This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
54	useConstraints This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
55	orderProcess This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
56	agencyProgram This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
57	sourcesOfData This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
58	methodology This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
59	supplementalInformation This element may include the element wellKnown and may also include locally defined elements.	Constructed as follows --
70	availability This element may include any of the following as well as locally defined elements: distributor, resourceDescription, orderProcess, technicalPrerequisites, timePeriod, linkage, linkageType.	Constructed as follows --
71	spatialReference This element may include any of the following as well as locally defined elements: boundingRectangle, geographicName.	Constructed as follows --
90	distributor This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumer, faxNumber.	Constructed as follows --
91	boundingRectangle This element may include any of the following as well as locally defined elements: westernMost, easternMost, northernMost, southernMost.	Constructed as follows --

GILS tagSet

<u>Tag</u>	<u>Element</u>	<u>Recommended Datatype</u>
92	geographicName	Constructed as follows -- This element may include any of the following as well as locally defined elements: geographicKeywordName, geographicKeywordType.
93	timePeriod	Constructed as follows -- This element may include any of the following as well as locally defined elements: timePeriodStructured, timePeriodTextual.
94	pointOfContact	Constructed as follows -- This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumber, faxNumber.
95	controlledVocabulary	Constructed as follows -- This element may include any of the following as well as locally defined elements: indexTermsControlled, thesaurus.
96	indexTermsControlled	Constructed as follows -- This element may include the following as well as locally defined elements: controlledTerm.
97	localSubjectIndex	Constructed as follows -- This element may include the following as well as locally defined elements: localSubjectTerm.
98	crossReference	Constructed as follows -- This element may include any of the following as well as locally defined elements: title, linkage, linkageType.

GILS Abstract Record Structure

NOTE: The element "bodyOfDisplay" in tagSet-G (2,9) may be used by the target to combine into this single element (i.e., bodyOfDisplay) one or more of the elements from the following abstract record structure into a display format.

<u>Tag</u>	<u>Element</u>	<u>Mandatory?</u>	<u>Repeatable?</u>
<u>path</u>			
(1,10)	rank	N	N
(1,12)	url	N	N
(1,14)	local control number	Y	N
(1,16)	dateOfLastModification	Y	N
(4,50)	title	Y	N
(4,1)	controlIdentifier	Y	N
(2,6)	abstract	Y	N

<u>Tag</u>	<u>Element</u>	<u>Mandatory?</u>	<u>Repeatable?</u>
<u>path</u>			
(4,51)	purpose	Y	N
(4,52)	originator	Y	N
(4,53)	accessConstraints	Y	N
(4,54)	useConstraints	Y	N
(4,70)	availability	Y	Y
(4,70)/(4,90)	distributor	Y	N
(4,70)/(4,90)/(2,7)	distributorName	Y	N
(4,70)/(4,90)/(2,10)	distributorOrganization	Y	N
(4,70)/(4,90)/(4,2)	distributorStreetAddress	Y	N
(4,70)/(4,90)/(4,3)	distributorCity	Y	N
(4,70)/(4,90)/(4,4)	distributorState	Y	N
(4,70)/(4,90)/(4,5)	distributorZipCode	Y	N
(4,70)/(4,90)/(2,16)	distributorCountry	Y	N
(4,70)/(4,90)/(2,12)	distributorNetworkAddress	Y	Y
(4,70)/(4,90)/(4,6)	distributorHoursofService	Y	Y
(4,70)/(4,90)/(2,14)	distributorPhoneNumber	Y	Y
(4,70)/(4,90)/(2,15)	distributorFaxNumber	Y	Y
(4,70)/(4,7)	resourceDescription	N	N
(4,70)/(4,55)	orderProcess	Y	N
(4,70)/(4,8)	technicalPrerequisites	N	N
(4,70)/(4,93)	timePeriod	N	Y
(4,70)/(4,93)/(4,15)	timePeriodStructured	N	Y
(4,70)/(4,93)/(4,16)	timePeriodTextual	N	Y
(4,70)/(4,17)	linkage	N	N
(4,70)/(4,18)	linkageType	N	N
(4,94)	pointOfContact	Y	N
(4,94)/(2,7)	contactName	Y	N
(4,94)/(2,10)	contactOrganization	Y	N
(4,94)/(4,2)	contactStreetAddress	Y	N
(4,94)/(4,3)	contactCity	Y	N
(4,94)/(4,4)	contactState	Y	N
(4,94)/(4,5)	contactZipCode	Y	N
(4,94)/(2,16)	contactCountry	Y	N
(4,94)/(2,12)	contactNetworkAddress	Y	Y
(4,94)/(4,6)	contactHoursofService	Y	Y
(4,94)/(2,14)	contactPhoneNumber	Y	Y
(4,94)/(2,15)	contactFaxNumber	Y	Y
(4,19)	recordSource	Y	N
(4,56)	agencyProgram	N	N
(4,57)	sourcesOfData	N	N

<u>Tag</u>	<u>Element</u>	<u>Mandatory?</u>	<u>Repeatable?</u>
<u>path</u>			
(4,95)	controlledVocabulary	N	Y
(4,95)/(4,96)	indexTermsControlled	Y	N
(4,95)/(4,96)/(4,20)	controlledTerm	Y	Y
(4,95)/(4,21)	thesaurus	Y	N
(4,97)	localSubjectIndex	N	N
(4,97)/(4,22)	localSubjectTerm	Y	Y
(4,58)	methodology	N	N
(4,71)	spatialReference	N	N
(4,71)/(4,91)	boundingRectangle	N	N
(4,71)/(4,91)/(4,9)	westernMost	N	N
(4,71)/(4,91)/(4,10)	easternMost	N	N
(4,71)/(4,91)/(4,11)	northernMost	N	N
(4,71)/(4,91)/(4,12)	southernMost	N	N
(4,71)/(4,92)	geographicName	N	Y
(4,71)/(4,92)/(4,13)	geographicKeywordName	Y	N
(4,71)/(4,92)/(4,14)	geographicKeywordType	Y	N
(4,93)	timePeriod	N	Y
(4,93)/(4,15)	timePeriodStructured	N	N
(4,93)/(4,16)	timePeriodTextual	N	N
(4,98)	crossReference	N	Y
(4,98)/(4,50)	crossReferenceTitle	Y	N
(4,98)/(4,17)	crossReferenceLinkage	Y	N
(4,98)/(4,18)	crossReferenceType	Y	N
(4,23)	originalControlIdentifier	Y	N
(4,59)	supplementalInformation	Y	N

Annex E

GILS Core Elements

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The term "mandatory" as used in this Profile applies to administration of the subset of GILS Locator Records that have been identified by the record source as participating in the GILS Core. GILS servers are not required to distinguish "mandatory" from other elements.

TITLE (Mandatory, Not Repeatable): This element conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject.

CONTROL IDENTIFIER (Mandatory, Not Repeatable): This element is defined by the information provider and is used to distinguish this locator record from all other GILS Core locator records. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual.

ABSTRACT (Mandatory, Not Repeatable): This element presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length.

PURPOSE (Mandatory, Not Repeatable): This element describes why the information resource is offered and identifies other programs, projects, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources.

ORIGINATOR (Mandatory, Not Repeatable): This element identifies the information resource originator, named as in the U.S. Government Manual where applicable.

ACCESS CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any access constraints applied to assure the protection of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here.

USE CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any use constraints applied to assure the protection of privacy or intellectual property and any other special restrictions or limitations on using the information resource.

AVAILABILITY (Mandatory, Repeatable): This element is a grouping of subelements that together describe how the information resource is made available.

DISTRIBUTOR (Mandatory, Not Repeatable): This subelement consists of the following subordinate fields that provide information about the distributor:

- DISTRIBUTOR NAME
- DISTRIBUTOR ORGANIZATION
- DISTRIBUTOR STREET ADDRESS
- DISTRIBUTOR CITY
- DISTRIBUTOR STATE
- DISTRIBUTOR ZIP CODE
- DISTRIBUTOR COUNTRY
- DISTRIBUTOR NETWORK ADDRESS
- DISTRIBUTOR HOURS OF SERVICE
- DISTRIBUTOR TELEPHONE
- DISTRIBUTOR FAX

RESOURCE DESCRIPTION (Optional, Not Repeatable): This subelement identifies the resource as it is known to the distributor.

ORDER PROCESS (Mandatory, Not Repeatable): This subelement provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods.

TECHNICAL PREREQUISITES (Optional, Not Repeatable): This subelement describes any technical prerequisites for use of the information resource as made available by this distributor.

AVAILABLE TIME PERIOD (Optional, Repeatable): This subelement provides the time period reference for the information resource as made available by this distributor, in one of two forms:

TIME PERIOD -- STRUCTURED: Time described using the USMARC prescribed structure.

TIME PERIOD -- TEXTUAL: Time described textually.

AVAILABLE LINKAGE (Optional, Not Repeatable): This subelement provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information.

AVAILABLE LINKAGE TYPE (Optional, Not Repeatable): This subelement occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI.

POINT OF CONTACT FOR FURTHER INFORMATION (Mandatory, Not Repeatable): This element identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact. This element consists of the following subelements:

CONTACT NAME
CONTACT ORGANIZATION
CONTACT STREET ADDRESS
CONTACT CITY
CONTACT STATE
CONTACT ZIP CODE
CONTACT COUNTRY
CONTACT NETWORK ADDRESS
CONTACT HOURS OF SERVICE
CONTACT TELEPHONE
CONTACT FAX

RECORD SOURCE (Mandatory, Not Repeatable): This element identifies the organization, as named in the U.S. Government Manual, that created or last modified this locator record.

DATE OF LAST MODIFICATION (Mandatory, Not Repeatable): This element identifies the latest date on which this locator record was created or modified.

AGENCY PROGRAM (*, Not Repeatable): This element identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource.

* This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

SOURCES OF DATA (*, Not Repeatable): This element identifies the primary sources or providers of data to the system, whether within or outside the agency.

* This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

CONTROLLED VOCABULARY (Optional, Repeatable): This element is a grouping of subelements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary:

INDEX TERMS -- CONTROLLED (Optional, Not Repeatable): This subelement is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field:

CONTROLLED TERM.

THESAURUS (Optional, Not Repeatable): This subelement provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. Notes on how to obtain electronic access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.

LOCAL SUBJECT INDEX (Optional, Not Repeatable): This element is a grouping of descriptive terms to aid users in locating resources of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating subelement:

LOCAL SUBJECT TERM

METHODOLOGY (Optional, Not Repeatable): This element identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described.

SPATIAL REFERENCE (Optional, Not Repeatable): This element is a grouping of subelements that together provide the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

BOUNDING RECTANGLE (Optional, Not Repeatable): This subelement provides the limits of coverage expressed by latitude and longitude values in the order:

WESTERN-MOST
EASTERN-MOST
NORTHERN-MOST
SOUTHERN-MOST.

GEOGRAPHIC NAME (Optional, Repeatable): This subelement identifies significant areas and/or places within the coverage through two associated constructs:

GEOGRAPHIC KEYWORD NAME
GEOGRAPHIC KEYWORD TYPE.

TIME PERIOD OF CONTENT (Optional, Repeatable): This element provides time frames associated with the information resource, in one of two forms:

TIME PERIOD – STRUCTURED: Time described using the USMARC prescribed structure.

TIME PERIOD – TEXTUAL: Time described textually.

CROSS REFERENCE (Optional, Repeatable): This element is a grouping of subelements that together identify another locator record likely to be of interest:

CROSS REFERENCE TITLE (Mandatory, Not Repeatable): This subelement provides a human readable textual description of the cross reference.

CROSS REFERENCE LINKAGE (Mandatory, Not Repeatable): This subelement provides the machine readable information needed to perform the access (i.e., URI).

CROSS REFERENCE TYPE (Mandatory, Not Repeatable): This subelement occurs if there is a CROSS REFERENCE LINKAGE and provides the data content type (i.e., MIME) for the referenced URI.

ORIGINAL CONTROL IDENTIFIER (Optional, Not Repeatable): This element is used by the record source to refer to another GILS locator record from which this locator record was derived.

SUPPLEMENTAL INFORMATION (Optional, Not Repeatable): Through this element, the record source may associate other descriptive information with the GILS Core locator record.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork.				
1. AGENCY USE ONLY (leave blank)		2. REPORT DATE October 1994		3. REPORT TYPE AND DATES COVERED Conference Publication Sept. 21-22, 1994
4. TITLE AND SUBTITLE CENDI Indexing Workshop			5. FUNDING NUMBERS	
6. AUTHOR(S)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Headquarters			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546			10. SPONSORING / MONITORING AGENCY REPORT NUMBER NASA CP-10160	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Unclassified—Unlimited Subject Category 82			12b. DISTRIBUTION CODE	
13. ABSTRACT The CENDI Indexing Workshop held at NASA Headquarters, Two Independence Square, 300 E street, Washington, DC, on September 21-22, 1994 focused on the following topics: machine aided indexing, indexing quality, an indexing pilot project, the MedIndEx Prototype, Department of Energy/Office of Scientific and Technical Information indexing activities, high-tech coding structures, category indexing schemes, and the Government Information Locator Service. This publication consists mostly of viewgraphs related to the above noted topics. In an appendix is a description of the Government Information Locator Service.				
14. SUBJECT TERMS Indexes (documentation), terms, information theory, probability theory, information retrieval, data processing, conferences, document storage, information dissemination, conferences, libraries			15. NUMBER OF PAGES 305	
			16. PRICE CODE A 13	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

Available from NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
(301) 621-0390